

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



**Estudo da região controlo total do DNA mitocondrial
numa população de imigrantes oriundos de Moçambique
integrados na população de Lisboa**

Marta Maria Proença de Campos

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Mestre Heloísa Afonso Costa
Professor Doutor Manuel do Carmo Gomes

2017

“O que prevemos raramente ocorre;
o que menos esperamos geralmente acontece.”

Benjamin Disraeli

Agradecimentos

As nossas realizações pessoais e profissionais dependem, grande parte, daqueles que caminham ao nosso lado, bem como, com quem nos cruzamos ao longo do nosso percurso. Como tal, não podia deixar de agradecer a todos os que contribuíram, direta e indiretamente, para a concretização e finalização de mais uma etapa da minha vida.

Em primeiro lugar, à minha orientadora Dra. Heloísa Afonso Costa, quero deixar o meu mais profundo agradecimento. Obrigada por ter aceite a orientação da presente Dissertação e por ter confiado em mim ao longo deste percurso. Obrigada, ainda, por todo o apoio incondicional, pela alegria contagiante, amizade, motivação e paciência demonstradas desde o primeiro dia.

Ao meu orientador Dr. António Amorim, por me ter dado a oportunidade de participar e realizar este meu projeto. Um agradecimento especial por toda a disponibilidade e acompanhamento diário demonstrados, pela paciência, ensinamentos, amizade e incentivos constantes. Obrigada por ter confiado em mim!

Ao Professor Dr. Manuel do Carmo Gomes, por ter aceite a orientação da presente Dissertação, na qualidade de meu orientador interno, e por estar sempre disponível no esclarecimento de dúvidas que surgiram ao longo da realização da tese.

À Dra. Teresa Ribeiro, responsável pelo Serviço de Genética e Biologia Forenses na Delegação Sul, pelo acolhimento e integração e por facilitar a concretização deste trabalho.

À Dra. Cláudia Vieira da Silva, um obrigada pelas longas conversas à hora do almoço e pelo incentivo e amizade demonstrada desde o início.

À restante equipa do Serviço de Genética e Biologia Forenses da Delegação Sul, um muito obrigada por me terem recebido bem e pela ajuda prestada.

A ti, Soraia, por teres caminhado comigo no desenvolvimento deste projeto e pela amizade e momentos que passamos juntas. Sem dúvida que tornaste todo este percurso mais fácil.

Às minhas melhores amigas, Rita, Sara e Sofia, os meus mais sinceros agradecimentos. A conclusão e realização desta etapa não seria possível sem a vossa amizade. Foram imprescindíveis desde o começo desta aventura. Obrigada por acreditarem e confiarem sempre em mim. À Filipa, um especial muito obrigada por teres estado presente quando mais precisava. Agradeço-te pela motivação, incentivo e ajuda que me proporcionaste durante o desenrolar deste projeto. Ao Nelson, ao Rui e ao Bruno um obrigada pelo apoio, paciência e amizade que sempre demonstraram.

Um agradecimento especial também aos meus queridos CELindos pela amizade, partilha e alegria demonstrada desde o início. A ti, Ana Belinha, o meu mais sincero obrigada. Agradeço-te por teres sido o meu porto de abrigo e por me teres dado a mão sempre que eu precisava. Foste incansável comigo. À Daniela, Irma, Gladys, Cheila, Patrícia, Cláudia, Mário e Luís obrigada por terem contribuído, cada um à sua maneira, para a realização da presente Dissertação.

Ao Rafa, por me aturar nos momentos de desespero, quando o meu computador não queria colaborar com a escrita desta Dissertação.

Ao Mário, o meu mais sincero Obrigada. Desculpa a falta de tempo, as ausências e a impaciência que demonstrei ao longo do percurso da tese. Obrigada por toda a compreensão, ajuda, amor e carinho que me proporcionaste durante este tempo. Sem ti, o término desta etapa não teria sido possível. Obrigada por estares sempre comigo!

Por fim, mas não menos importante, um reconhecimento muito especial ao meu irmão, pais e avós por terem sempre confiado em mim e apoiado as minhas escolhas e decisões. Obrigada por todo o esforço, incentivo incondicional, motivação e dedicação ao longo desta aventura. Sem a vossa preciosa ajuda jamais teria chegado até aqui! Obrigada por me terem ensinado a lutar e a não desistir.

A ti, mãe, devo-te tudo. Não há palavras para te agradecer tudo o que fizeste por mim. Espero conseguir, um dia, retribuir tudo o que tens feito por mim. Obrigada!

A todos, os meus mais sinceros agradecimentos.

Resumo

O número de imigrantes a residir em Portugal tem vindo a aumentar exponencialmente desde o final dos anos 70. Portugal passou a ser um destino desejado principalmente por indivíduos provenientes de países de língua oficial portuguesa. De acordo com dados estatísticos divulgados até ao final de 2015, a população portuguesa era constituída, aproximadamente, por 400 000 imigrantes, dos quais 3 000 eram oriundos de Moçambique. Destes 3 000, cerca de 1 700 residiam na Região Metropolitana de Lisboa.

O estudo do DNAm_t tem-se tornado uma ferramenta essencial nas mais diversas áreas do conhecimento. Desde a antropologia, à genética populacional e forense, tem sido considerado um marcador genético de eleição, devido às suas características peculiares, como o elevado número de cópias por célula, a hereditariedade uniparental materna, a elevada taxa de mutação e a ausência de recombinação. Estas propriedades do DNAm_t permitem reconstruir a origem e evolução das populações, bem como, investigar o seu passado demográfico e a dinâmica, de modo a traçar a história de uma determinada população e a estabelecer uma correlação com eventos da humanidade.

Dada a sua relevância, o objetivo principal deste trabalho prende-se com a caracterização genética da população imigrante moçambicana que reside, atualmente, na Grande Lisboa, com a finalidade de averiguar a potencial diversidade genética que estes indivíduos vêm introduzir na população de Lisboa e, consequentemente, em Portugal.

Para a realização do estudo, um total de 83 indivíduos imigrantes moçambicanos foram analisados. A região controlo do DNAm_t foi amplificada e sequenciada na sua totalidade, utilizando dois pares de *primers* - L15971/H016 e L16555/H639. O conjunto de haplótipos obtidos foi submetido e aceite para integrar a base de dados internacional de DNAm_t mais conceituada no âmbito das Ciências Forenses, a EMPOP, com o código de acesso EMP00681.

Aproximadamente 82% dos haplótipos observados neste estudo revelaram ser únicos, reforçando a grande variabilidade genética das populações africanas. A maioria das sequências de DNAm_t identificadas (81%) correspondem a haplogrupos característicos e típicos de regiões subsarianas: o macrohaplogrupo L. A partir dos estudos filogenéticos comparativos verificou-se que a população de imigrantes moçambicanos se encontra geneticamente mais próxima da população de imigrantes angolanos a residir em Lisboa e geneticamente mais distante da população portuguesa.

Palavras-chave: DNAm_t; Região controlo; Haplogrupos; Genética populacional; Imigrantes moçambicanos.

Abstract

Since the late 70's the number of immigrants in Portugal has exponentially increase. Portugal has been chosen as a living destiny, mostly by people from countries with Portuguese as an official language. According to the last 2015 statistics, Portugal has approximately 400 000 immigrants, of which 3 000 are from Mozambique. From those 3 000, about 1 700 live in Lisbon.

MtDNA has become an essential tool in many fields. Since anthropology to the forensic and populational genetics, it has been selected as the main genetic biomarker due to its peculiar characteristics such as the high number of copies per cell, the uniparental inheritance from the mother, the high mutation rate and the absence of recombination. These properties allow to reconstruct the origin and evolution of a population, as well as investigate their demographic past and population dynamics, in order to establish the history of a certain population and its correlation with humanity events.

Given its relevance, the main goal of this work is to characterize genetically the Mozambican population who lives in Lisbon, and try to inquire the potential genetic diversity that these people may introduce in the capital's population and in Portugal.

For this study, 83 samples from Mozambican immigrants were analysed. The mtDNA control region was amplified and sequenced using two pairs of primers - L15971/H016 and L16555/H639. The set of obtained haplotypes was submitted and accepted to be a part in the most prestigious international database of mtDNA – EMPOP – with the access code EMP00681.

Approximately 82% of the observed haplotypes in this study showed to be unique, enhancing the genetic variability in the African populations. Most of the mtDNA regions identified (81%) correspond to typical haplotypes from the Subsharian region: macrohaplogroup L. Taking in account the comparative phylogenetic studies, we conclude that the Mozambican immigrant population is genetically closer to the Angolan immigrant population living in Lisbon and further away from the Portuguese population.

Key words: mtDNA; Control region; Haplogroups; Population genetics; Mozambican immigrants.

Comunicações Científicas

Posters

M. Proença de Campos, H. Afonso Costa, C. Vieira da Silva, T. Ribeiro, M. J. Porto, A. Amorim (2016) **Genetic portrait of Lisboa immigrant population from Mozambique with mitochondrial DNA**. Haploid Markers 2016. Berlim.

Marta Proença de Campos, Heloísa Afonso Costa, Cláudia Vieira da Silva, Vanessa Bogas, Teresa Ribeiro, Maria João Porto, **António Amorim** (2017) **The immigrant population from Mozambique in Lisbon: updated mitochondrial DNA portrait**. ISFG 2017. Seul.

Artigo Científico

Marta Proença de Campos, Heloísa Afonso Costa, Cláudia Vieira da Silva, Vanessa Bogas, Teresa Ribeiro, Maria João Porto, António Amorim, “**The immigrant population from Mozambique in Lisbon: updated mitochondrial DNA portrait**,” *Forensic Science International: Genetics Supplement Series*, 2017. DOI: 10.1016/j.fsigss.2017.09.134

Índice

Agradecimentos	III
Resumo	V
Abstract	VI
Comunicações Científicas	VII
Índice de Figuras	X
Índice de Tabelas	XI
Lista de Abreviaturas e Siglas	XII

1. Introdução	1
1.1. Mitocôndria	1
1.2. Genoma Mitocondrial	1
1.2.1. Características do DNAm	2
1.2.2. Heteroplasmia	3
1.2.3. Aplicações do DNAm	4
1.2.3.1. Genética Populacional	4
1.2.3.2. Aplicações Forenses	5
1.3. Bases de Dados	6
1.4. Migração e População	6
2. Objetivos	7
3. Materiais e Métodos	8
3.1. Seleção das Amostras	8
3.2. Extração de DNA	8
3.3. Amplificação do DNA	8
3.4. Purificação dos Produtos Amplificados	10
3.5. Sequenciação	10
3.6. Purificação dos Produtos Sequenciados	11
3.7. Detecção dos Produtos Sequenciados	11
3.8. Análise dos Produtos Sequenciados	12
3.9. Análise Filogenética	13
4. Resultados e Discussão	14

4.1. Estudo dos Haplótipos	14
4.2. Estudo das Heteroplasmias	19
4.3. Estudo dos Haplogrupos	20
4.3.1. Haplogrupo L0	21
4.3.2. Haplogrupo L1	22
4.3.3. Haplogrupo L2	22
4.3.4. Haplogrupo L3	23
4.3.5. Haplogrupo L5	23
4.3.6. Haplogrupos euroasiáticos	24
4.4. Análise Filogenética	25
5. Conclusões	28
6. Estudos Futuros	29
7. Referências	30
8. Anexos	40
8.1. Anexo A	40

Índice de Figuras

Figura 1.1 – Representação esquemática da região controlo e da região codificante da molécula de DNAm (adaptado de [9]). 1

Figura 1.2 – Mapa representativo da expansão do Homem pelo mundo e respetiva distribuição geográfica dos haplogrupos (adaptado de [50]). 5

Figura 3.1 – Exemplo de um segmento de um eletroferograma obtido no programa *Sequencing Analysis* v.5.2, após sequenciação com o *primer* L16555. 11

Figura 3.2 – Representação, através de um eletroferograma, de um segmento da região HVI (a) com heteroplasma de comprimento e (b) com heteroplasma de posição. Ambas as amostras foram sequenciadas com o *primer* L15971 e estão identificadas pela seta cor-de-laranja. 13

Figura 4.1 – Representação dos diferentes haplogrupos e da respetiva frequência relativos aos 83 indivíduos da população imigrante moçambicana. 21

Figura 4.2 – Representação gráfica das distâncias genéticas entre a população imigrante moçambicana e as restantes selecionadas da literatura. 26

Índice de Tabelas

Tabela 3.1 – Sequências dos <i>primers</i> utilizados durante o processo de amplificação da região controlo total do DNA mitocondrial.	9
Tabela 3.2 – Reagentes e respetivos volumes utilizados, por amostra, na preparação da reação de amplificação da região controlo total do DNA mitocondrial.	9
Tabela 3.3 – Condições de temperaturas e tempos requeridos para a amplificação da região controlo total do DNA mitocondrial.	9
Tabela 3.4 - Reagentes e respetivos volumes utilizados, por amostra, na preparação da reação de sequenciação da região controlo total do DNA mitocondrial.	10
Tabela 3.5 - Condições de temperaturas e tempos requeridos para a sequenciação da região controlo total do DNA mitocondrial.	11
Tabela 3.6 - Código IUPAC utilizado na nomenclatura das bases nucleotídicas do DNAm.	12
Tabela 4.1 – Haplótipos mais frequentes observados no presente estudo e respetiva frequência.	15
Tabela 4.2 – Distribuição das alterações polimórficas identificadas nas 83 amostras da população imigrante moçambicana, por tipo de polimorfismo e por segmento hipervariável da região controlo total do DNAm.	15
Tabela 4.3 – Polimorfismos observados em baixa frequência, na região HVI do DNAm.	16
Tabela 4.4 – Alterações polimórficas presentes em baixa frequência, na região HVII do DNAm. ...	17
Tabela 4.5 – Substituições observadas em baixa frequência, na região HVIII do DNAm.	18
Tabela 4.6 – Distribuição das amostras heteroplásmicas por tipo de heteroplasma e por segmento hipervariável da região controlo do DNAm identificadas nas 83 sequências analisadas.	19
Tabela 4.7 – Populações selecionadas da literatura para estudos comparativos filogenéticos, tendo por base a região controlo total do DNAm.	25
Tabela 4.8 – Valores de F_{st} e p relativos às populações do estudo filogenético obtidos através do programa <i>Arlequin</i>	25
Tabela 8.1 – Conjunto dos resultados obtidos para os 83 indivíduos imigrantes moçambicanos, quanto ao número de alterações polimórficas determinadas relativamente à rCRS, haplótipo e respetivo haplogrupo e codificação atribuída a cada sequência.	42

Lista de Abreviaturas e Siglas

A – Adenina

AB – Applied Biosystems

DNA – Ácido desoxirribonucleico

DNAm – Ácido desoxirribonucleico mitocondrial

DNAn – Ácido desoxirribonucleico nuclear

AMOVA – Análise da variância molecular

ATP – Adenosina trifosfato

C – Citosina

CRS – Sequência de referência de Cambridge

d – Deleção

ddATP – Didesoxiadenosina-trifosfato

ddCTP – Didesoxicitosina-trifosfato

ddGTP – Didesoxiguanina-trifosfato

ddNTPs – Didesoxinucleotídeos trifosfatados

ddTTP – Didesoxitimidina-trifosfato

DEL – Deleção

dNTPs – Desoxinucleotídeos trifosfatados

EDNAP – *European DNA Profiling Group*

EMPOP – *EDNAP mitochondrial DNA population database*

F_{st} – Índice de fixação

G – Guanina

H – Cadeia pesada

HV – Região hipervariável

HVI – Segmento hipervariável 1

HVII – Segmento hipervariável 2

HVIII – Segmento hipervariável 3

InDels – Inserção/Deleção

INMLCF – DS – Instituto Nacional de Medicina Legal e Ciências Forenses - Delegação Sul

IUPAC – *International Union of Pure & Applied Chemistry*

L – Cadeia leve

MgCl₂ – Cloreto de Magnésio

O_H – Origem de replicação da cadeia pesada

p – Nível de significância

PCR – *Polymerase Chain Reaction*

poli-C – Poli-citosinas

R – Adenina e guanina

rCRS – Sequência de referência de Cambridge revista

RFLP – *Restriction fragment length polymorphism*

ROS – Espécies reativas de Oxigênio

rRNAs – Ácido ribonucleico ribossomal

rSAP – Fosfatase alcalina recombinante de camarão

SGBF – Serviço de Genética e Biologia Forenses

SNPs – *Single Nucleotide Polymorphism*

STRs – *Short Tandem Repeats*

T – Timina

Trans – Transições

tRNAs – Ácido ribonucleico de transferência

UEPs – *Unique Event Polymorphisms*

VNTRs – *Variable Number of Tandem Repeats*

Y – Citosina e timina

adeninas -, de maior peso molecular, e por uma cadeia leve (L – *Light*), complementar à cadeia pesada, com menor peso molecular, rica em pirimidinas - citosinas e timinas [10,11].

Relativamente à função, o DNAm pode ser dividido em duas regiões: uma região codificante e uma região não-codificante. A região codificante constitui a maior extensão da molécula e codifica um conjunto de 37 genes, intervenientes no processo de fosforilação oxidativa e produção de energia, que incluem 13 genes de polipéptidos, 22 de tRNAs e 2 de rRNAs [7]. A região não-codificante, também designada por região controlo ou *D-loop*, compreende cerca de 7% do genoma mitocondrial e abrange um total de 1 122 pares de base (bp), que se estende desde a posição 16 024 à 16 569 e da 1 à 576 [12,13]. É nesta última região que se pode identificar a origem de replicação da cadeia pesada (O_H), os promotores de transcrição de ambas as cadeias e os locais de ligação de fatores de transcrição [14]. A região controlo do DNAm é composta por segmentos hipervariáveis (HV – *Hypervariable region*), altamente polimórficos, considerados a fração mais heterogénea da molécula, e nos quais se pode observar a maior parte dos polimorfismos identificados no genoma mitocondrial. São três os segmentos hipervariáveis descritos: HVI compreendido entre as posições 16024 à 16365, HVII que se estende da posição 73 à 340 e HVIII entre as posições 438 à 574. É nas regiões HVI e HVII que se verifica a maior densidade de alterações polimórficas [14,15].

A sequência completa do genoma mitocondrial, com 16 569 bp, foi determinada e publicada pela primeira vez em 1981, por S. Anderson [16]. Esta sequência, considerada a sequência de referência de estudos que recorrem ao DNAm, ficou conhecida como Sequência de Referência de Cambridge (CRS – *Cambridge Reference Sequence*). No seu estudo, S. Anderson estabeleceu que a numeração dos nucleótidos tinha início na região controlo da molécula, numa posição próxima à origem de replicação da cadeia pesada [16]. Em 1999, R. Andrews observou incoerências na CRS [17]. Resequenciou e reviu a mesma amostra utilizada por S. Anderson, o que permitiu a identificação e correção de 11 erros [17]. Assim, a sequência corrigida foi publicada e passou a ser designada por Sequência de Referência de Cambridge revista (rCRS – *revised Cambridge Reference Sequence*), sequência esta que tem sido utilizada até à atualidade [18]. A referida sequência pertence a um indivíduo que se enquadra num haplogrupo tipicamente europeu, o haplogrupo H, e tem em conta a cadeia leve da molécula de DNAm.

Quando se compara a sequência de uma amostra com a rCRS, apenas as diferenças existentes são apresentadas. O conjunto de alterações observado na sequência de DNAm de um indivíduo é designado por haplótipo e é descrito com o número da posição mutada e com a base nucleotídica que originou a alteração. Os haplótipos, depois, vão ser agrupados em haplogrupos. Um haplogrupo consiste num grupo de haplótipos que partilham características comuns e, como tal, considera-se que descendem de um mesmo antecessor comum [19,20]. A sua classificação é efetuada com letras maiúsculas de A a Z.

1.2.1. Características do DNAm

O DNAm possui características que o tornam um marcador genético de eleição, principalmente em estudos de genética populacional e genética forense. São elas o elevado número de cópias, a hereditariedade materna, a elevada taxa de mutação e a ausência de recombinação.

O genoma mitocondrial, ao contrário do ácido desoxirribonucleico nuclear (DNAn), que contém apenas duas cópias por célula, apresenta um elevado número de cópias por célula humana. Este número varia consoante o tipo de célula ou tecido, oscilando entre 2 a 10 cópias de DNA por mitocôndria, que corresponde, aproximadamente, a 10^3 a 10^4 cópias por célula [12,21,22]. O elevado número de cópias é, assim, uma propriedade do DNAm que confere vantagem quando a quantidade de DNAn é baixa ou quando está altamente degradado nas amostras.

Outra característica importante do DNAm é a hereditariedade materna. Em 1980, R. Giles chegou à conclusão, que o DNAm era herdado exclusivamente ou principalmente por via materna [23]. Neste processo, as mitocôndrias paternas e, por consequência, o DNAm paterno são, segundo alguns autores, destruídas e eliminadas seletivamente no oócito, durante a fertilização, por um mecanismo de ubiquitinação [24] ou por diluição [20]. A transmissão uniparental faz do DNAm um sistema genético haplóide extremamente útil na identificação de membros de uma família que partilhem o mesmo antecessor materno, visto que os seus genomas mitocondriais são idênticos [20,25].

Devido à sua elevada taxa de mutação, o DNAm é uma molécula extremamente polimórfica, que evolui 5 a 10 vezes mais rapidamente do que o genoma nuclear [26]. Esta é outra das características que faz do DNAm um bom marcador genético. A taxa de mutação alcançada para o genoma mitocondrial varia entre $0,075 \times 10^{-6}$ e $0,165 \times 10^{-6}$ substituições/posição/ano, consoante o local do genoma onde as alterações ocorrem [27]. A região controlo do DNAm apresenta uma taxa de mutação bastante heterogênea, englobando locais onde se observa maior incidência de alterações. Esses locais, designados de *hotspots* mutacionais, exibem uma evolução 4 a 5 vezes mais rápida comparado com as restantes posições da mesma região [27]. São algumas as razões que tentam explicar a alta taxa de mutação do DNAm. O fato desta molécula ser vulnerável à falta de histonas protetoras e ao ataque oxidativo provocado por espécies reativas de oxigénio (ROS), libertadas pela cadeia transportadora de eletrões, faz com que o DNAm seja suscetível à acumulação de mutações [1,22]. Os mecanismos de reparação de erros do DNAm são também menos eficientes que os mecanismos nucleares devido à baixa fidelidade da polimerase do DNAm, levando, por consequência, a mutações [22,28]. Um dos tipos de alterações que se pode observar no DNAm são as substituições. Estas, comparativamente com as inserções e deleções, apresentam, na região controlo, uma incidência maior [12]. As substituições podem ser classificadas em transições, que ocorrem entre pirimidinas ou entre purinas, e transversões, no qual uma pirimidina é substituída por uma purina ou vice-versa. As transições, mais especificamente as transições entre pirimidinas, são as alterações mais prevalentes no DNAm [15].

A aparente falta de recombinação da molécula de DNAm é outra das características que fazem deste genoma um bom marcador genético. Contudo, esta particularidade é bastante controversa entre diferentes investigadores. Enquanto uns referem uma possível recombinação da molécula [29–31], outros não detetaram qualquer evidência [32–34]. De fato, sabe-se que a molécula de DNAm é detentora de um sistema funcional capaz de efetuar recombinação homóloga [35]. No entanto, a taxa em que esta pode ocorrer é extremamente ínfima, pelo que não contribui para a variação genética deste genoma [35,36]. A variabilidade encontrada, assim, na molécula de DNAm é consequência do ganho de mutações adquiridas ao longo de várias gerações [27]. Esta ausência de recombinação permite estabelecer relações filogenéticas entre indivíduos que partilhem a mesma linhagem materna, sendo, por isso, muitas vezes utilizada em casos de investigação de identidade [20].

1.2.2. Heteroplasmia

As heteroplasmas, quando detetadas, podem ser consideradas uma característica peculiar do DNAm. Define-se heteroplasmia como a coexistência de duas ou mais populações de moléculas de DNAm que diferem, ligeiramente, na sua composição nucleotídica, numa única mitocôndria, célula ou indivíduo [37,38]. Estas foram identificadas, pela primeira vez, nos anos 80, em estudos genéticos de doente com patologia mitocondrial [39]. Sabe-se, hoje, que as heteroplasmas são bastante mais frequentes do que se pensava antigamente, podendo ser identificadas em qualquer indivíduo da população em geral [40].

As heteroplasmas podem ser classificadas em heteroplasmas de posição ou de comprimento. As heteroplasmas de posição são geralmente mutações pontuais numa das sequências de um indivíduo, resultando na mistura de nucleótidos numa mesma posição. Na leitura do eletroferograma será visível

a sobreposição de dois nucleótidos diferentes [41,42]. As heteroplasmias de comprimento, por sua vez, estão associadas a regiões polimórficas constituídas por citosinas, designadas de zonas homopoliméricas de citosinas ou de poli-citosinas (poli-C) ou *hotspots mutacionais* [43]. Ocorrem tipicamente em sequências que apresentam, no mínimo, 8 citosinas consecutivas. É na região controlo do DNAm, nomeadamente nos segmentos HVI e HVII, que se observa a maior frequência das heteroplasmias de comprimento [41]. A sua manifestação na região codificante é rara e poderá estar associada a patologias [44,45]. O eletroferograma irá apresentar múltiplos picos que diferem, em comprimento, num único nucleótido, tornando, assim, difícil a análise do genoma [46–48].

Inúmeros estudos realizados ao DNAm sugerem que o padrão de heteroplasmia não é uniforme no próprio indivíduo [49]. De fato, sabe-se que os níveis de heteroplasmia podem variar consideravelmente entre células do mesmo tecido ou órgão e entre diferentes tecidos dentro de um mesmo indivíduo e/ou entre indivíduos da mesma linhagem materna [50,51]. O tecido muscular esquelético e o cerebral são os tecidos que apresentam os índices mais elevados de heteroplasmia [41].

Há vários aspetos que ainda permanecem controversos e sob investigação no que se refere a heteroplasmias no DNAm. O local onde ocorrem, a probabilidade de serem herdadas e a provável variação com a idade, género, grupo populacional e tipo de tecido precisam de ser ainda esclarecidos [43].

1.2.3. Aplicações do DNAm

Nos últimos 30 anos, o genoma mitocondrial tem sido considerado uma ferramenta poderosa nas mais diversas áreas. Desde a antropologia molecular e arqueologia, à genética médica, genética populacional e evolutiva e genética forense, tem sido um marcador genético de eleição, devido às suas características peculiares [27,52].

Os estudos efetuados inicialmente à molécula de DNAm basearam-se em técnicas de restrição enzimática, de baixa resolução e baixa capacidade de discriminação, que utilizavam a totalidade da molécula de DNAm, técnica esta designada por RFLP (*Restriction fragment length polymorphism*) [23,27,53]. Com a evolução das metodologias de análise, nomeadamente da técnica PCR (*Polymerase Chain Reaction*) e da sequenciação de Sanger, e com o estudo mais pormenorizado do genoma mitocondrial, o aumento da qualidade e credibilidade dos resultados era incontestável. Os primeiros estudos de sequenciação da molécula de DNAm exploravam apenas o segmento HVI [54,55] ou os segmentos HVI e HVII da região controlo [56,57], uma vez que nestes dois segmentos hipervariáveis se observa a maior taxa de mutação e, por consequência, a grande parte dos polimorfismos identificados [12]. Mais recentemente, o estudo do genoma mitocondrial foi alargado ao segmento HVIII da região controlo e à totalidade da região controlo do DNAm [58,59], o que aumenta significativamente o poder de individualização entre amostras. Com as tecnologias de nova geração, a introdução do estudo da região codificante do DNAm [60,61] garante uma eficácia superior na discriminação entre indivíduos e na análise dos polimorfismos e atribuição dos haplogrupos.

1.2.3.1. Genética Populacional

Devido às características particulares do DNAm, muitos estudos têm sido realizados com o objetivo de reconstruir a origem e evolução das populações e de investigar o passado demográfico e a dinâmica populacional, nomeadamente, das expansões, migrações e outros padrões de fluxo de genes. Desta forma é possível traçar a história de uma determinada população e estabelecer uma correlação com eventos da humanidade [27,62,63].

A teoria mais aceite, atualmente, relativamente à origem do homem moderno é a hipótese *Out of Africa*, na qual se preconiza que o homem teve uma origem africana, a partir da qual se terá dispersado para o resto do mundo (figura 1.2) [13,64,65]. A análise das linhagens maternas de DNAm permite,

assim, retirar e adicionar informações acerca da origem do homem moderno e da sua dispersão. Após inúmeros estudos, estima-se que o ancestral feminino comum surgiu há aproximadamente 200 000 anos, em África, tendo divergido e transmitido a sequência de DNAm_t à população atual [65,66]. Este antepassado feminino é designado por Eva mitocondrial [67].

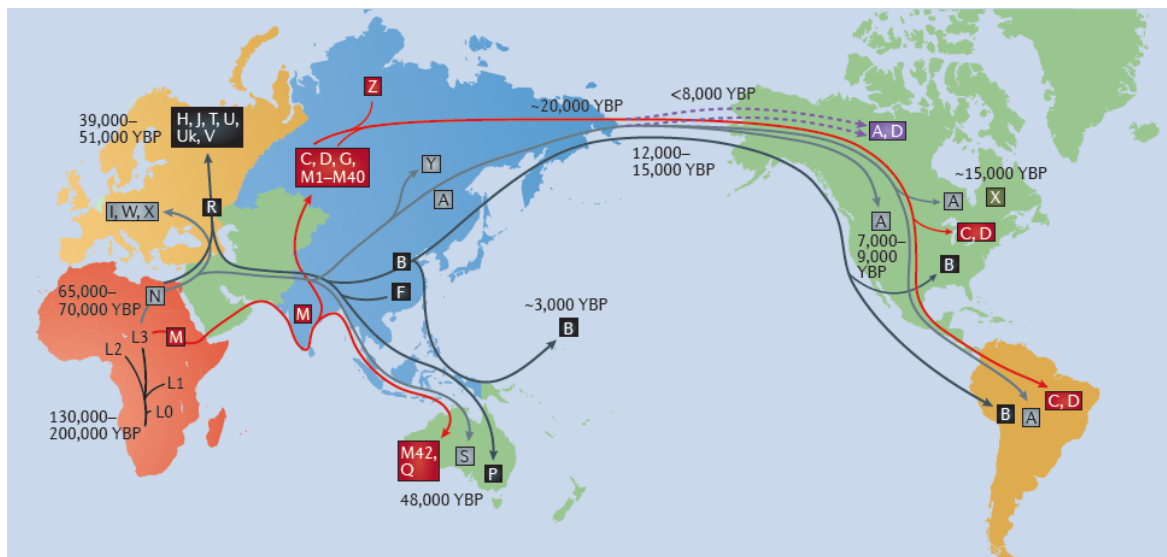


Figura 1.2 – Mapa representativo da expansão do Homem pelo mundo e respetiva distribuição geográfica dos haplogrupos (adaptado de [50]).

A evolução de uma dada população é refletida, na sequência do DNAm_t, pela acumulação de mutações. Estas mutações definem os haplótipos que, por sua vez, se agrupam em haplogrupos [27]. Os haplogrupos de DNAm_t podem estar associados a determinadas regiões geográficas e/ou a grupos étnicos em particular.

O conjunto das alterações observadas nos indivíduos é organizado e utilizado para construir árvores filogenéticas [68], sendo a *PhyloTree* a árvore filogenética mais utilizada atualmente [52]. Esta árvore contém polimorfismos quer da região controlo quer da região codificante do DNAm_t, estando permanentemente a ser atualizada. A base da árvore filogenética engloba todos os haplogrupos característicos de populações africanas, nomeadamente de regiões de África subsariana, o macrohaplogrupo L [27,66,69], que se ramifica nos haplogrupos L0, L1, L2, L3, L4, L5 e L6 [52]. Foi a partir deste macrohaplogrupo L, principalmente do haplogrupo L3, que os restantes haplogrupos divergiram. Este haplogrupo L3, dispersou para regiões europeias e asiáticas, dando origem aos macrohaplogrupos M e N, que caracterizam as linhagens presentes nestas regiões. O macrohaplogrupo N, típico de populações europeias, divergiu nos haplogrupos H, I, J, K, T, U, V, W e X, que constituem aproximadamente 99% dos haplogrupos desta região [20,27,70]. As sequências de DNAm_t, características de regiões asiáticas, englobam haplogrupos que pertencem tanto ao macrohaplogrupo M como ao N [71]. De todas as populações analisadas, as que apresentam maior variabilidade e heterogeneidade genética são as populações do continente africano [72]. Em média, as alterações nucleotídicas destas populações são aproximadamente o dobro das variações observadas nos indivíduos euroasiáticos [73].

1.2.3.2. Aplicações Forenses

A utilização de métodos analíticos, nomeadamente SNPs (*Single Nucleotide Polymorphism*), VNTRs (*Variable Number of Tandem Repeats*), STRs (*Short Tandem Repeats*) e sequenciação direta

têm ajudado na resolução de casos de investigação forense [20,74]. Quando o DNAn das amostras biológicas está presente em quantidade e qualidade desejáveis, estes métodos revelam ser eficazes e praticamente infalíveis [20,74]. Apesar de revelar menor poder de discriminação comparativamente com o DNAn, o DNAm tem demonstrado ser uma mais valia em análises de amostras cuja quantidade de DNA extraído é muito reduzida ou na presença de amostras altamente degradadas [12,46,74]. Tal fato é justificado pelas propriedades únicas do genoma mitocondrial, bem como pela forma circular da molécula de DNAm, que a torna menos vulnerável à ação de enzimas exonucleases, sendo, por isso, mais resistente ao processo de degradação [75]. A utilidade, aplicação e validação do DNAm, nesta área, estão bem documentadas na literatura [76,77].

A utilização do DNAm em genética forense é vantajosa em situações em que o objetivo é comparar indivíduos que pertencem a uma mesma linhagem materna, dado que as sequências de DNAm destes indivíduos são similares. Esta característica pode ser particularmente útil na identificação de um familiar desaparecido [20]. A análise do DNAm também apresenta desvantagens. A principal é o baixo poder de discriminação e individualização. Dois indivíduos que partilhem uma mesma linhagem materna apresentam o mesmo haplótipo, salvo raras mutações, não sendo possível fazer distinção entre eles [78,79].

Uma das mais conhecidas investigações nesta área, efetuada em restos cadavéricos, é o da família de Nicholas II Romanov, último Czar da Rússia [80,81].

1.3. Bases de Dados

Com o aumento exponencial de estudos com o DNAm, surgiu a necessidade de criar bases de dados mitocondriais com o objetivo de melhorar a qualidade dos resultados publicados, ao minimizar erros decorrentes de metodologias e análises, e de padronizar dados, de modo a ser possível efetuar comparações [82,83]. Estas bases de dados são constituídas por haplótipos a partir dos quais é possível realizar comparações, estimar incidências de haplótipos e determinar frequências de mutações numa determinada população [82]. Atualmente, são inúmeras as bases de dados especializadas que se encontram descritas na literatura, destacando-se a EMPOP [84], o MITOMAP [85], o mtDB [86], a MitoVariome [87], o MitBase [88] e o MitoBreak [89].

A EMPOP – *EDNAP mitochondrial DNA population database* é a base de dados internacional mais conceituada e utilizada em estudos de DNAm. Esta base de dados consiste num projeto, acessível a toda a comunidade científica, que pretende aumentar a qualidade e confiança das sequências de DNAm de diversas populações [84,90,91]. Todos os haplótipos presentes na EMPOP são submetidos a um rigoroso controlo de qualidade previamente à publicação, para diminuir erros de análise e metodologia e aumentar a qualidade das sequências disponíveis [90]. Em 2007, a EMPOP armazenava um total de 5173 haplótipos [84]. Dez anos mais tarde, em 2017, 34 617 haplótipos estão alojados nesta base de dados (www.empop.org).

1.4. Migração e População

Portugal, embora considerado um país tradicionalmente de emigração, tem, desde os finais dos anos 70, observado um intenso fluxo de imigrantes, nomeadamente de indivíduos pertencentes a países de língua oficial portuguesa, incluindo Moçambique. Sendo a migração um evento que aumenta consideravelmente a heterogeneidade genética, este movimento populacional, ao longo de Portugal, vai desempenhar um importante papel na modulação dos padrões de diversidade genética, tornando crucial os estudos genéticos destas populações de imigrantes [92,93].

Até finais de 2016, Portugal era constituído, aproximadamente, por 400 000 imigrantes, dos quais 3 000 eram oriundos de Moçambique. Destes 3 000, cerca de 1 700 residiam na Região Metropolitana de Lisboa [94].

2. Objetivos

O objetivo principal deste trabalho é a realização de um estudo genético, da população imigrante oriunda de Moçambique a residir na Grande Lisboa, através da caracterização da região controlo total do DNAm dos referidos imigrantes. Esta caracterização tem como finalidade averiguar a diversidade e heterogeneidade genética e o potencial impacto que estes indivíduos vêm introduzir na população de Lisboa, e consequentemente, em Portugal. Neste sentido, delinear-se os seguintes objetivos específicos:

- Determinar e definir os haplótipos de DNAm para cada indivíduo;
- Determinar os haplogrupos correspondentes a cada haplótipo;
- Inferir, a partir de cada haplogrupo e subhaplogrupo, a origem geográfica provável de cada indivíduo;
- Submeter os haplótipos e respetivos haplogrupos à base de dados internacional de DNAm mais importante e relevante na área das Ciências Forenses, a EMPOP - *EDNAP Forensic mtDNA Population Database*, de modo a disponibilizar os resultados a toda a comunidade científica;
- Determinar a distância genética entre a população de imigrantes em estudo e outras populações africanas, selecionadas da literatura, bem como a distância genética destas populações em relação à população portuguesa da região Sul;
- Elaborar uma comparação interpopulacional, de modo a esclarecer as suas relações genéticas.

3. Materiais e Métodos

3.1. Seleção das Amostras

Neste estudo foram analisadas 83 amostras de sangue de imigrantes moçambicanos, não relacionados entre si, que residem atualmente na região de Lisboa. As amostras estudadas eram provenientes de investigações de parentesco biológico, que decorreram no Serviço de Genética e Biologia Forenses (SGBF) da Delegação Sul do Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF – DS), entre 2002 e 2015. Das amostras selecionadas, 34 pertenciam a indivíduos do género masculino e 49 a indivíduos do género feminino. Todas as amostras utilizadas receberam previamente uma codificação com o objetivo de impedir a associação a cada indivíduo.

3.2. Extração de DNA

A extração de DNA a partir de manchas de sangue foi realizada utilizando o método do Chelex®100 (Bio-Rad), uma resina quelante que impede a degradação do DNA. Esta resina liga iões magnésio que, sendo cofatores de enzimas nucleases, tornam as enzimas inativas, protegendo assim o DNA da destruição. Esta técnica, desenvolvida em 1991 por P. Sean Walsh [95], consiste num conjunto de várias etapas descritas de seguida.

Para cada amostra foi recortado um círculo de mancha de sangue, armazenada em cartão *Whatman™ FTA™*, com recurso a um *paper punch single hole*. Colocou-se o fragmento num *eppendorf* de 1,5 mililitro (mL), devidamente identificado, e adicionou-se 1 mL de água mili-Q. Incubaram-se os tubos à temperatura ambiente, durante cerca de 20 minutos e, de seguida, fez-se uma agitação ligeira no vórtex, durante 5 a 10 segundos. As amostras foram centrifugadas a 14 000 rotações por minuto (rpm), durante 3 minutos, a 20°C. Com auxílio de uma micropipeta, e evitando o contato com o *pellet*, que contém o material genético, retirou-se o sobrenadante de cada tubo, aproximadamente 980 µL, e adicionou-se 180 µL de Chelex®100 a 5%. Esta solução tinha sido previamente preparada por diluição de 2,5 g de Chelex®100 em 50 mL de água mili-Q. De seguida, colocaram-se os tubos num banho de água aquecido, a 56°C, durante 15 minutos. Posteriormente, os tubos foram sujeitos a nova agitação no vórtex, a alta velocidade, durante cerca de 15 segundos, e colocados em suportes flutuantes. Estes suportes com as amostras foram inseridos num banho de água em ebulição, a 100°C, durante 8 minutos. Este passo é crucial, pois é aqui que efetivamente ocorre a extração do DNA. Por fim, os tubos foram novamente agitados no vórtex, a alta velocidade, e centrifugados a 14 000 rpm, durante 5 minutos, a 20°C. As amostras foram armazenadas a -20°C, temperatura esta que previne a ação das enzimas nucleases, até posterior utilização. Em cada processo de extração foi preparado um controlo negativo, sujeito às condições anteriormente descritas, ao qual se adicionou água mili-Q em substituição de amostra de sangue.

3.3. Amplificação do DNA

A amplificação do DNA tem como objetivo o aumento exponencial do número de cópias de uma região específica do DNA que se pretende estudar. A amplificação é alcançada pela adição de *primers* específicos e complementares às extremidades 3' de cada uma das sequências que delimitam o fragmento pretendido [96].

Para a amplificação da região controlo total do DNAm, constituída pelas regiões HVI, HVII e HVIII, utilizaram-se dois pares de *primers*. O par L15971 (direto) e H016 (reverso) amplificou a região entre as posições 16024 e a 16569 enquanto que o par L16555 (direto) e H639 (reverso) amplificou a região entre as posições 1 a 576. A nomenclatura dos *primers* é definida de acordo com a sua afinidade para a cadeia leve (L) ou pesada (H) bem como pela posição nucleotídica na região 3' com a qual interagem. A sequência específica de cada *primer* encontra-se representada na tabela 3.1.

Tabela 3.1 – Sequências dos *primers* utilizados durante o processo de amplificação da região controlo total do DNA mitocondrial.

<i>Primers</i>	Sequência
L15971	5' - TTA ACT CCA CCA TTA GCA CC - 3'
H016	5' - TGA TAG ACC TGT GAT CCA TCG TGA - 3'
L16555	5' - CCC ACA CGT TCC CCT TAA AT - 3'
H639	5' - GGG TGA TGT GAG CCC GTC TA - 3'

Para a realização desta técnica foram utilizados vários reagentes: *Multiplex PCR Master Mix*[®] (Quiagen) que contém tampão de reação com Cloreto de Magnésio (MgCl₂), *HotStarTaq Plus DNA polymerase* e desoxinucleotídeos trifosfatados (dNTPs), o conjunto de *primers* anteriormente mencionados e o DNA extraído de cada uma das amostras, num volume total de reação de 10 µL. Os volumes necessários de cada reagente encontram-se descritos na tabela 3.2.

Tabela 3.2 – Reagentes e respetivos volumes utilizados, por amostra, na preparação da reação de amplificação da região controlo total do DNA mitocondrial.

Reagentes	Volume
<i>Multiplex PCR Master Mix</i>	5 µL
Água mili-Q	3 µL
<i>Primers</i> a 2 µM	1 µL
DNA	1 µL
Total	10 µL

Para cada reação de amplificação prepararam-se dois controlos negativos, um para cada par de *primers*, aos quais se adicionou água mili-Q em substituição de DNA. A reação de amplificação da região controlo total foi efetuada num termociclador *GeneAmp PCR system 9700* (Applied Biosystems - AB) programado de acordo com a tabela 3.3, num total de 35 ciclos.

Tabela 3.3 – Condições de temperaturas e tempos requeridos para a amplificação da região controlo total do DNA mitocondrial.

1 ciclo	35 ciclos			1 ciclo	
Incubação Inicial	Desnaturação	Annealing	Extensão	Extensão Final	Hold
94°C 5 min	94°C 30 seg	60°C 90 seg	72°C 60 seg	72°C 10 min	4°C ∞

3.4. Purificação dos Produtos Amplificados

Este passo de purificação foi realizado com recurso ao método do *ExoSAP-IT*[®] (Affymetrix) e teve como objetivo a remoção de elementos em excesso provenientes da reação de amplificação. Este reagente contém duas enzimas hidrolíticas – exonuclease I e fosfatase alcalina recombinante de camarão (rSAP) - que são responsáveis pela remoção, respetivamente, de *primers* residuais e de dNTPs não incorporados no decorrer do processo de amplificação.

A técnica de purificação consistiu na adição de 4 µL de *ExoSAP-IT*[®] por cada 10 µL de produto amplificado, num volume final de 14 µL. Esta etapa foi programada e realizada num termociclador *GeneAmp PCR system 9700* (AB). Inicialmente incubaram-se as amostras a 37°C, temperatura definida como ótima para as enzimas que constituem o reagente, durante 15 minutos. Seguiu-se um período de aquecimento a 80°C, também durante 15 minutos, com o intuito de inativar as enzimas do reagente.

3.5. Sequenciação

A sequenciação da região controlo do DNAm_t tem por objetivo determinar as bases nucleotídicas características de cada posição dentro desta região. O método utilizado nesta etapa foi o método de sequenciação de Sanger [97], que consiste na incorporação de didesoxinucleotídeos trifosfatados (ddNTPs) na extremidade 3' da cadeia molde, levando ao término da síntese da cadeia na posição onde ocorre a ligação dos ddNTPs. Estes ddNTPs apresentam na sua constituição marcadores fluorescentes, com diferentes cores, que permitem, durante a análise das sequências, a sua deteção e identificação. O ddATP encontra-se marcado a verde, o ddCTP a azul, o ddGTP marcado a amarelo mas visualizado a preto, e o ddTTP marcado a vermelho.

Nesta fase do processo foi necessária a preparação de quatro reações, uma por cada *primer* (L15971, L16555, H016 e H639). Para além dos *primers* da cadeia a amplificar, a mistura era constituída por tampão *Better Buffer*[®] (Microzone Ltd.), *BigDye*[®] *Terminator v.3.1 Cycle Sequence* (AB), uma solução que contém DNA polimerase, MgCl₂, dNTPs e ddNTPs marcados com *Dye Terminator*, num volume total de 4 µL. As soluções foram distribuídas pelos poços de uma placa de 96 poços, aos quais se adicionou 1 µL de DNA das amostras previamente amplificadas e purificadas. Os volumes necessários de cada reagente encontram-se descritos na tabela 3.4.

Tabela 3.4 - Reagentes e respetivos volumes utilizados, por amostra, na preparação da reação de sequenciação da região controlo total do DNA mitocondrial.

Reagentes	Volume
<i>Better Buffer</i>	2 µL
<i>BigDye Terminator</i>	1 µL
<i>Primers</i> a 2.5 µM	1 µL
DNA	1 µL
Total	5 µL

Prepararam-se simultaneamente quatro controlos negativos, um por cada *primer*, aos quais se adicionou água mili-Q em substituição de DNA amplificado e purificado. O processo de sequenciação da região controlo total do DNAm_t foi realizado num termociclador *GeneAmp PCR system 9700* (AB) programado para as temperaturas e os tempos descritos na tabela 3.5, num total de 35 ciclos.

Tabela 3.5 - Condições de temperaturas e tempos requeridos para a sequenciação da região controlo total do DNA mitocondrial.

1 ciclo	35 ciclos			1 ciclo	
Incubação Inicial	Desnaturação	Annealing	Extensão	Extensão Final	Hold
96°C 2 min	96°C 15 seg	50°C 9 seg	60°C 2 min	60°C 10 min	4°C ∞

3.6. Purificação dos Produtos Sequenciados

A etapa de purificação foi realizada com *BigDye® XTerminator Purification Kit* (AB), que possui dois reagentes na sua composição, o *XTerminator Solution* e a *SAM TM Solution*. Este passo é importante na remoção do excesso de dNTPs e ddNTPs não incorporados durante a reação de sequenciação, bem como na remoção do excesso de *primers* e sais livres.

Neste processo preparou-se uma solução de reação constituída por 30 µL de água, 16,4 µL de *SAM Solution* e 3,7 µL de *XTerminator Solution*, num volume final de 50,1 µL por amostra. Distribuíram-se 50 µL da mistura por cada poço utilizado da placa de 96 poços. Colocou-se a placa num agitador, durante aproximadamente 30 minutos. Por fim, efetuou-se uma centrifugação a 2800 rpm, durante 2 minutos, passo que permitiu a deposição dos excessos da reação de sequenciação, impedindo a sua interferência na eletroforese. Desta forma, durante a eletroforese capilar, os capilares aspiram apenas os produtos sequenciados retidos no sobrenadante.

3.7. Detecção dos Produtos sequenciados

Após purificação dos produtos sequenciados, as amostras foram colocadas num sequenciador automático, o *Genetic Analyser 3130* (AB) de 4 capilares, para serem submetidas a uma eletroforese capilar. A realização desta eletroforese possibilitou a deteção dos fragmentos obtidos e a aquisição da sequência nucleotídica da região controlo do DNAm em estudo. Durante este processo, os fragmentos migram, pelos capilares, consoante o seu peso molecular, desde o polo negativo (cátodo) até ao polo positivo (ânodo). Isto significa que fragmentos de tamanho menor irão migrar a uma velocidade maior do que fragmentos de tamanho maior. Aquando da migração, um feixe laser incide sobre os fragmentos que, ao excitar os eletrões presentes nos fluorocromos dos ddNTPs, promovem a emissão de fluorescência num determinado comprimento de onda. Os resultados obtidos de cada amostra por cada *primer* organizam-se num eletroferograma (figura 3.1), que constitui, através de picos, o conjunto de todos os nucleótidos que compõem a região controlo.

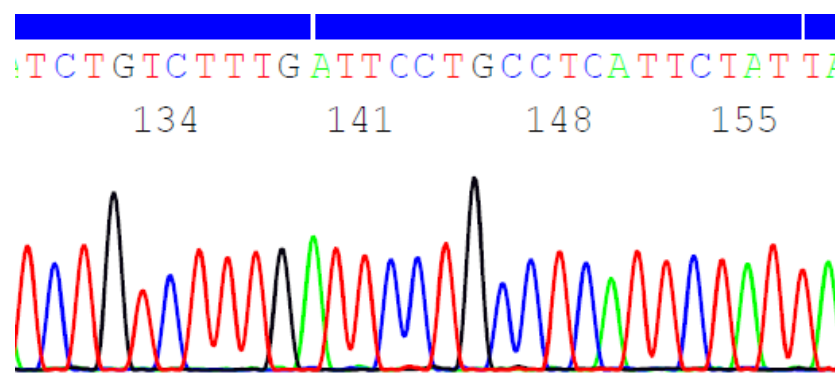


Figura 3.1 – Exemplo de um segmento de um eletroferograma obtido no programa *Sequencing Analysis v.5.2*, após sequenciação com o *primer* L16555.

Previamente ao início da corrida e com auxílio do programa 3130 *Series Data Collection Software* v.4, as amostras e os controlos presentes em cada um dos poços da placa de 96 poços foram identificados com a codificação da amostra e o *primer* correspondente.

3.8. Análise dos produtos Sequenciados

A região controlo total do DNAm, compreendida entre as posições 16024 a 16569 e 1 a 576, foi sequenciada em todas as amostras dos indivíduos em estudo. Ambos os intervalos mencionados foram sequenciados utilizando um par de *primers* (um direto e outro reverso), em que um tem afinidade para a cadeia leve e o outro para a cadeia pesada. O objetivo deste procedimento é comparar os segmentos sequenciados por cada par de *primers*, de forma a verificar se há concordância, em cada posição nucleotídica, entre as cadeias sequenciadas pelos *primers* direto e reverso.

A análise dos resultados da sequenciação da região controlo do DNAm foi efetuada com auxílio a dois programas, o *Sequencing Analysis* v.5.2, que avalia a qualidade das sequências por identificação dos nucleótidos em cada posição, e o *SeqScape* v.3 (AB), que permitiu o alinhamento e a posterior comparação das sequências obtidas com a rCRS [17]. Este último programa identifica apenas as diferenças observadas entre a sequência em estudo e a sequência padrão, discriminando, assim, para cada amostra, todas as bases nucleotídicas e respetivas posições que se encontrem alteradas. Para cada indivíduo, o conjunto de alterações observadas relativamente à rCRS dá origem ao haplótipo. A caracterização das sequências teve por base as *guidelines* da *European DNA Profiling Group* (EDNAP) [98] e as regras de nomenclatura da *International Union of Pure & Applied Chemistry* (IUPAC) [99]: A (adenina), C (citocina), G (guanina), T (timina), R (adenina e guanina) e Y (citocina e timina) (tabela 3.6).

Tabela 3.6 - Código IUPAC utilizado na nomenclatura das bases nucleotídicas do DNAm.

Código IUPAC	Base(s) nucleotídica(s)	Código IUPAC	Base(s) nucleotídica(s)
A	Adenina	Y	Citocina ou Timina
C	Citocina	K	Guanina ou Timina
G	Guanina	M	Adenina ou Citocina
T	Timina	N	Qualquer base
B	Citocina, Guanina ou Timina	V	Adenina, Citocina ou Guanina
D	Adenina, Guanina ou Timina	S	Guanina ou Citocina
H	Adenina, Citocina ou Timina	W	Adenina ou Timina
R	Adenina ou Guanina		

A região controlo total do DNAm é uma região hipervariável que pode acumular alterações nucleotídicas. As alterações por substituição podem ser classificadas em transições, que podem ocorrer entre purinas (A e G) ou entre pirimidinas (C e T), e transversões, que ocorrem entre uma purina e uma pirimidina ou vice-versa. Para detetar substituições é necessário comparar as sequências em estudo com a rCRS, após a qual se identifica a posição em que ocorreu a alteração e o nucleótido que originou essa substituição. Outro tipo de alterações que podem ser encontradas durante a análise do DNAm são as inserções. Estas são identificadas pela posição nucleotídica imediatamente anterior à alteração, seguida de um ponto e do número um (.1) e do nucleótido inserido, para a primeira inserção, de um ponto e do número dois (.2) e do nucleótido inserido, na segunda inserção, e assim

sucessivamente, consoante o número de inserções. As deleções são outro tipo de alterações que podem ser determinados na análise do DNAm. Estas são representadas pela posição do nucleótido deletado relativamente à rCRS, seguida do sufixo “d”, “DEL” ou “-”.

Nas amostras que apresentaram heteroplasma de comprimento, a leitura da sequência só foi possível até à posição heteroplásmica, uma vez que, a partir desta região, o eletroferograma é uma mistura de bases nucleotídica sobrepostas e ilegíveis (figura 3.2a). Para estes casos, com o intuito de confirmar a sequência obtida, efetuou-se a repetição da sequenciação do segmento heteroplásmico. Outra característica que se observou em algumas das sequências de DNAm analisadas foi a presença de heteroplasma de posição, que consiste na sobreposição de nucleótidos numa determinada posição da sequência (figura 3.2b). A sua identificação segue as diretrizes descritas no código da IUPAC e representam-se pela posição heteroplásmica e a letra correspondente.

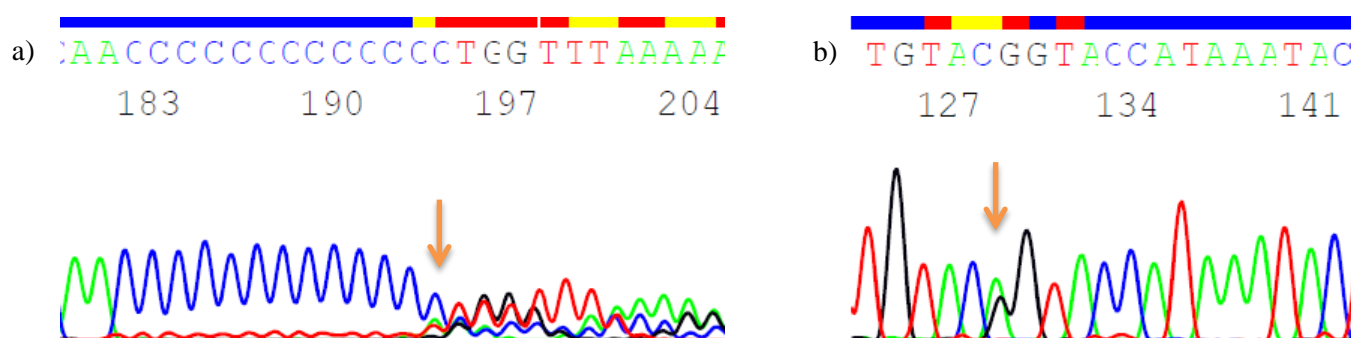


Figura 3.2 – Representação, através de um eletroferograma, de um segmento da região HVI (a) com heteroplasma de comprimento e (b) com heteroplasma de posição. Ambas as amostras foram sequenciadas com o *primer* L15971 e estão identificadas pela seta cor-de-laranja.

Para se proceder à classificação dos haplótipos em haplogrupos recorreu-se a uma árvore filogenética de DNAm, a *PhyloTree*, desenhada em 2009, por van Oven e Kayser [52]. Esta árvore exibe os polimorfismos existentes relativamente à rCRS.

3.9. Análise Filogenética

Para determinar as relações filogenéticas entre a população em estudo e as selecionadas da literatura, efetuou-se a análise da variância molecular (AMOVA). Este parâmetro permite calcular, através do estudo dos haplótipos, a variabilidade genética entre diferentes populações. Para realizar a AMOVA e estimar as distâncias genéticas interpopulacionais, utilizou-se o programa *Arlequin* v. 3.5.2.2 [100]. Através deste programa foram obtidos os valores do índice de fixação (F_{st}), que avalia as distâncias genéticas entre as populações, e o valor de p , que indica o nível de significância (0.05).

Tendo por base os valores de F_{st} determinados e recorrendo ao programa *Neighbor do software Phylip* v.3.695 [101] foi possível obter as distâncias moleculares entre as sequências de DNAm das populações analisadas. A sua representação gráfica, em forma de árvore filogenética, foi elaborada através do programa *TreView* v. 1.6.6 [102].

4. Resultados e Discussão

Para cumprir o principal objetivo deste estudo determinaram-se os haplótipos que caracterizam os 83 indivíduos em estudo e os respectivos haplogrupos. O conjunto de dados obtidos foi submetido a uma base de dados internacional, a EMPOP [84], e foi aceite para integrar a mesma, assegurando assim a qualidade das sequências e dos haplogrupos definidos. Estes estão disponíveis para a comunidade científica com o código de acesso EMP00681.

Até ao momento, relativamente à população moçambicana, não existem estudos na literatura referentes à totalidade da região controlo do DNAm. Publicações anteriores apenas reportam análises da região HVI [55] e da região HVI + HVII [56]. Em ambos os estudos, eliminaram-se regiões que apresentam polimorfismos característicos, já descritos, e que são de extrema importância no processo de discriminação. Este fato aumenta o número de haplótipos partilhados entre indivíduos, diminuindo o poder de individualização. O estudo da totalidade do DNAm torna-se, assim, vantajoso, uma vez que permite obter haplótipos mais pormenorizados e determinar haplogrupos com maior eficiência, levando, por consequência, ao aumento do poder de discriminação.

4.1. Estudo dos Haplótipos

Através dos dados obtidos neste estudo foi possível construir uma base de dados que contém, detalhadamente para cada amostra, todas as alterações encontradas, bem como a respetiva posição, o número de alterações existentes e o haplogrupo correspondente. O conjunto dos resultados está representado na tabela 8.1, em anexo.

Das 83 amostras analisadas identificaram-se 74 haplótipos diferentes, dos quais 68 (cerca de 82%) estavam presentes apenas uma única vez no conjunto das amostras estudadas. Sendo as populações do continente africano caracterizadas por uma grande variabilidade e heterogeneidade genética é comum observar uma elevada percentagem de haplótipos únicos [103]. Estudos anteriores de DNAm descrevem esta elevada frequência noutras populações de origem africana [104–109].

Os restantes 6 haplótipos identificados neste estudo foram partilhados por mais do que um indivíduo, num total de 15 sequências (18%) (tabela 4.1). Da análise dos resultados verificou-se que o haplótipo mais frequente estava presente em 4 das sequências (5%), sendo as alterações que o caracterizam descritas na tabela 4.1. Na rotina forense, a aplicação do DNAm em estudos de identificação individual depende da frequência dos haplótipos observados na população em geral [25]. O fato de haver haplótipos coincidentes entre indivíduos não relacionados demonstra que este marcador genético tem baixo poder de discriminação e individualização, não sendo determinante em investigações de identidade.

Tabela 4.1 – Haplótipos mais frequentes observados no presente estudo e respetiva frequência.

Amostras	Frequência	Haplótipos
Moc_050, Moc_083	2	16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
Moc_060, Moc_068	2	16223T 16278T 16286T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C
Moc_036, Moc_061	2	16223T 16278T 16286T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
Moc_002, Moc_047	2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523DEL 524DEL
Moc_009, Moc_020, Moc_057	3	16182C 16183C 16189C 16223T 16278T 16290T 16294T 16309G 16390A 73G 146C 152C 195C 263G 315.1C
Moc_003, Moc_042, Moc_051, Moc_070	4	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL

Comparando a região controlo total do DNAm das 83 sequências com a rCRS [17] identificaram-se 1295 alterações nucleotídicas, que correspondem a 127 posições polimórficas. Através da análise da tabela 4.2, verificou-se que as substituições foram o tipo de polimorfismo mais frequente, com 1055 alterações, que contrastam com as 84 deleções e as 151 inserções detetadas. Esta supremacia de substituições já foi reportada previamente por Carvalho para outras populações [110]. Observou-se ainda que 997 das substituições correspondem a transições e 58 a transversões, sendo que dentro das transições 641 ocorreram entre pirimidinas e 356 entre purinas, demonstrando-se assim que o tipo de transição predominante foi entre pirimidinas. Estes resultados foram consistentes com os obtidos por Budowle relativamente a amostras provenientes de vários continentes, incluindo do africano [73].

Tabela 4.2 – Distribuição das alterações polimórficas identificadas nas 83 amostras da população imigrante moçambicana, por tipo de polimorfismo e por segmento hipervariável da região controlo total do DNAm.

Alterações	HVI	HVII	HVIII	Fora HV	Total
Número total de Polimorfismos	533	559	114	89	1295
Número de Posições Polimórficas	67	36	16	8	127
Transições	480	417	12	88	997
Pirimidinas	393	178	9	61	641
C -> T	183	40	6	8	237
T -> C	210	138	3	53	404
Purinas	87	239	3	27	356
A -> G	62	187	3	1	253
G -> A	25	52	-	26	103
Transversões	44	13	-	1	58
A -> C	23	9	-	-	32
A -> T	3	-	-	-	3
C -> A	3	3	-	-	6
C -> G	14	-	-	-	14
G -> C	1	-	-	-	1
T -> A	-	1	-	1	2
Deleções	5	1	78	-	84
Inserções	-	127	24	-	151
Heteroplasmia de Posição	4	1	-	-	5
A e G (Y)	1	-	-	-	1
T e C (R)	3	1	-	-	4

A região HVI da região controlo do DNAm engloba os nucleótidos compreendidos entre as posições 16024 e 16365. Através da análise dos resultados, verificou-se que a posição 16223 apresentava em 67 das sequências analisadas (cerca de 81%) uma substituição de uma citosina (C) por uma timina (T), sendo esta alteração a mais comum em HVI. Observou-se também, em 39 das sequências estudadas (47%), a transição de uma citosina (C) por uma timina (T) na posição 16278. Em 36 das sequências obtidas (43%), identificou-se uma substituição de uma timina (T) por uma citosina (C) na posição 16311. Constatou-se ainda que a transição de uma citosina (C) por uma timina (T), na posição 16294, estava presente em 27 das sequências analisadas (32%).

O intervalo entre as posições 16182 e 16192 de HVI caracteriza-se pela sequência AACCCCTCCC. Esta zona de poli-C demonstrou uma elevada variabilidade nucleotídica, uma vez que só nesta região identificaram-se 116 polimorfismos, que correspondem a 9% de todas as alterações existentes. Na posição 16182 identificaram-se 9 sequências com mutação de uma adenina (A) por uma citosina (C). Na posição 16183 foram observadas 14 sequências com substituição de uma adenina (A) por uma citosina (C) e 2 sequências em que a adenina (A) se encontrava deletada. Foram também detetadas 2 sequências que apresentavam alteração de uma citosina (C) por uma timina (T), na posição 16185. Na posição 16187 verificou-se que 24 das sequências obtidas continham mutação de uma citosina (C) por uma timina (T). A posição 16188, constituída por uma citosina (C) na sequência de referência, encontra-se alterada em 17 das amostras estudadas, sendo que 14 das sequências apresentavam uma substituição para guanina (G), 2 estavam alteradas para adenina (A) e 1 apresentava transição para timina (T). Em 43 das sequências obtidas, na posição 16189 ocorreu uma mutação de timina (T) para citosina (C). Por último, verificou-se que, na posição 16192, 5 das sequências estavam alteradas de citosina (C) para timina (T). O conjunto dos polimorfismos 16187T 16188G 16189C foi observado em 14 das sequências, sendo o mais frequente nesta região.

Pela análise da tabela 8.1 foi possível detetar, na região HVI, uma série de polimorfismos que aparecem apenas uma vez nas 83 sequências obtidas. As 28 posições nucleotídicas (34%) que apresentam baixa frequência nas amostras estudadas encontram-se discriminadas na tabela 4.3.

Tabela 4.3 – Polimorfismos observados em baixa frequência, na região HVI do DNAm.

Posição	Trans Pirimidinas	Trans Purinas	Transversões	Posição	Trans Pirimidinas	Trans Purinas
16041		1 (A -->G)		16215		1 (A --> G)
16051		1 (A -->G)		16219		1 (A --> G)
16069	1 (C --> T)			16234	1 (C --> T)	
16092	1 (T --> C)			16261	1 (C --> T)	
16111	1 (C --> T)			16264	1 (C --> T)	
16114			1 (C --> A)	16266	1 (C --> T)	
16117	1 (T --> C)			16274		1 (G --> A)
16129			1 (G --> C)	16296	1 (C --> T)	
16145		1 (G --> A)		16304	1 (T --> C)	
16163		1 (A --> G)		16317		1 (A --> G)
16176	1 (C --> T)			16319		1 (G --> A)
16179	1 (C --> T)			16343		1 (A --> G)
16188	1 (C --> T)			16344	1 (C --> T)	
16213		1 (G --> A)		16357	1 (T --> C)	

Na tabela: Trans = Transições

A rCRS, na posição 16325 é constituída por uma timina (T). Através dos resultados obtidos, verificou-se que em 3 das sequências esta posição apresentava deleção da timina (T). No conjunto dos resultados obtidos para a região HVI constatou-se que não havia qualquer inserção.

No que diz respeito à região HVII da região controlo do DNAm, esta é constituída pelo conjunto de nucleótidos compreendidos entre as posições 73 e 340. A análise dos resultados mostrou que existe uma mutação extremamente comum, de adenina (A) para guanina (G), na posição 263, que estava presente em 77 das sequências analisadas (93%). A posição 73 apresentou, em 61 das sequências (73%), alteração de uma adenina (A) por uma guanina (G). Verificou-se ainda que a timina (T) da posição 152 foi substituída por uma citosina (C) em 47 das amostras (57%). Na posição 195 observou-se uma transição de timina (T) para citosina (C) em 36 das sequências (43%). Uma outra alteração identificada, relativamente à rCRS, foi a substituição de uma timina (T) por uma citosina (C), na posição 146, em 31 das sequências obtidas (37%).

Foi na região HVII que se observou a alteração mais comum deste estudo. Esta mutação foi identificada em todas as sequências analisadas e caracteriza-se pela inserção de uma citosina (C) na posição 315 (polimorfismo 315.1C). Esta inserção ocorre com frequência em comunidades africanas [106,107].

O segmento compreendido entre as posições 303 e 316 caracteriza-se pela sequência CCCCCCTCCCCCG. Nesta zona de poli-C, um total de 130 polimorfismos (10%) foram detetados. Através da tabela 8.1, verificou-se que, na posição 309, 36 das sequências apresentavam uma inserção de uma citosina (C; polimorfismo 309.1C) e que em 4 existia a inserção de duas citosinas (C; polimorfismos 309.1 e 309.2C). As inserções na zona de poli-C (posição 309 e 315) são muito frequentes em todos os grupos populacionais, incluindo nas populações africanas, fato já reportado em estudos anteriores [73]. Finalmente, em 3 das sequências obtidas observou-se ainda a substituição de uma guanina (G) por uma adenina (A), na posição 316.

Na região HVII verificou-se, tal como para HVI, a presença de polimorfismos que apresentavam baixa frequência, isto é, que estavam presentes uma única vez nas 83 sequências analisadas. Estas 13 posições nucleotídicas alteradas (16%) encontram-se descritas na tabela 4.4.

Tabela 4.4 – Alterações polimórficas presentes em baixa frequência, na região HVII do DNAm.

Posição	Trans Pirimidinas	Trans Purinas	Transversões
143		1 (G --> A)	
183		1 (A --> G)	
188		1 (A --> G)	
194	1 (C --> T)		
199	1 (T --> C)		
217	1 (T --> C)		
228		1 (G --> A)	
260		1 (G --> A)	
294			1 (T --> A)
295	1 (C --> T)		
297		1 (A --> G)	
325	1 (C --> T)		
340	1 (C --> T)		

Na tabela: Trans = Transições

A posição 249 na rCRS é representada por uma adenina (A). Numa das sequências obtidas observou-se deleção da adenina (A) que caracteriza esta posição. Para além das inserções já mencionadas, nenhuma outra foi detetada.

A região HVIII da região controlo do DNAm é formada pelo segmento compreendido entre as posições 438 e 574. A sequência que engloba as posições 515 a 524 caracteriza-se pela sequência ACACACACAC e integra um total de 80 polimorfismos (6%). Constatou-se, pela análise da tabela 8.1, que em 36 das sequências obtidas (43%), um dos dímeros AC estava deletado. Estas deleções, segundo a nomenclatura proposta por Carracedo [111], devem ser assinaladas como 523DEL e 524DEL, e têm uma elevada incidência em populações africanas, como demonstram estudos realizados à população do Gana e da Somália [106,107]. Foram também identificadas nesta região 2 sequências que apresentavam uma inserção dos dinucleótidos AC (polimorfismo 524.1A e 524.2C) e 1 sequência com inserção de dois conjuntos dos nucleótidos AC (polimorfismo 524.1A 524.2C 524.3A 524.4C).

O intervalo entre as posições 567 e 574 de HVIII caracteriza-se pela sequência ACCCCCCA. Nesta zona de poli-C foi possível verificar a existência de inserções, na posição 573, em 3 das sequências estudadas. Numa das sequências observou-se inserção de 3 citosinas (C; polimorfismo 573.1C 573.2C 573.3C). Noutra sequência foi detetada inserção de 4 citosinas (C; polimorfismo 573.1C 573.2C 573.3C 573.4C). Finalmente, constatou-se ainda a inserção de 5 citosinas (C) em outra das sequências analisadas (polimorfismo 573.1C 573.2C 573.3C 573.4C 573.5C).

Na região HVIII também foram observadas mutações que aparecem no máximo duas vezes no conjunto das sequências obtidas. Estas 10 posições nucleotídicas alteradas (12%) encontram-se descritas na tabela 4.5.

Tabela 4.5 – Substituições observadas em baixa frequência, na região HVIII do DNAm.

Posição	Trans Pirimidinas	Trans Purinas
444		2 (A --> G)
456	1 (C --> T)	
462	1 (C --> T)	
480	1 (T --> C)	
489	2 (T --> C)	
494	1 (C --> T)	
497	1 (C --> T)	
508		1 (A --> G)
511	1 (C --> T)	
527	1 (C --> T)	

Na tabela: Trans = Transições

Através da análise dos resultados, foi possível detetar, em 6 das sequências estudadas, a deleção de uma citosina (C) na posição 498. Verificou-se ainda a inserção de uma timina (T) na posição 455 e a inserção de uma citosina (C) na posição 459, partilhadas pelas mesmas 2 sequências.

Os segmentos da região controlo do DNAm compreendidos entre as posições 1 a 72, 341 a 437 e 16365 a 16569 não constituem as regiões hipervariáveis mencionadas anteriormente. No entanto, nas sequências analisadas, observaram-se substituições de nucleótidos nestes segmentos, num total de 89 (7%). Na posição 16519, 50 das sequências (60%) apresentavam alteração de uma timina (T) por uma

citossina (C). Em 26 das amostras, identificou-se na posição 16390 uma mutação de guanina (G) para adenina (A). A posição 64, que apresenta uma citossina (C) na rCRS, sofreu uma transição para timina (T) em 7 das sequências. Na posição 16368 detetou-se uma substituição de uma timina (T) por uma citossina (C) em 2 das sequências. As restantes 4 posições nucleotídicas mudadas - C375T, T398C, T408A, A16399G - aparecem apenas uma única vez nas sequências analisadas.

Ao comparar as sequências obtidas com a rCRS verificou-se que nenhuma das sequências estudadas apresentou total concordância. O menor número de alterações polimórficas visualizado nas sequências estudadas foi 5, num total de 3 sequências. Ao invés, a sequência que apresentou um maior número de polimorfismos contém um total de 30 substituições, caracterizando-se pelo conjunto 16117C 16129A 16172C 16173T 16188A 16189C 16223T 16256T 16278T 16291T 16293G 16294T 16311C 16360T 16368C 16519C 73G 151T 152C 182T 186A 195C 198T 247A 263G 297G 315.1C 316A 523DEL 524DEL. O número médio de alterações nucleotídicas encontradas relativamente à rCRS foi de 16 mutações por amostra. Para populações de origem africana a média de polimorfismos encontrados, segundo Budowle, foi 15, o que significa que o valor obtido no presente estudo é semelhante ao que já foi descrito anteriormente [73].

Com recurso ao *software Arlequin* foi possível obter o índice de diversidade de sequência, probabilidade de dois haplótipos escolhidos ao acaso diferirem, e o índice de diversidade nucleotídica, número de diferenças entre uma sequência e a rCRS. Os valores determinados foram, respetivamente, 0,9962 e 0,0152. A elevada diversidade encontrada neste estudo é concordante com estudos efetuados noutras populações do continente africano e evidencia a heterogeneidade característica destas populações quando comparadas com outras [106,108,109]. De realçar que os valores de diversidade calculados neste estudo são praticamente idênticos aos da população imigrante angolana [109]. Pereira, em 2001, para a região HVI + HVII (posição 16024 à 16365 e 73 à 340) determinou valores de diversidade de sequência de 0.973 [56]. Em 2002, Salas obteve valores de diversidade de sequência de 0.957, para a região HVI (16024 à 16365) [55]. Todos os valores mencionados, embora não comparáveis diretamente entre si, demonstram o poder de discriminação e individualização do estudo de diferentes regiões do DNAm. Quanto mais regiões, maior o poder de discriminação.

4.2. Estudo das Heteroplasmias

Nas 83 amostras analisadas, observou-se heteroplasmia em 23 das sequências, que corresponde a 28% das sequências estudadas (tabela 4.6). Valores semelhantes foram obtidos para outras populações do continente africano [106–109].

Tabela 4.6 – Distribuição das amostras heteroplásmicas por tipo de heteroplasmia e por segmento hipervariável da região controlo do DNAm identificadas nas 83 sequências analisadas.

Amostras	Tipo de Heteroplasmia	Segmento HV	Prevalência
Homoplásmicas	-	-	60
Heteroplásmicas			23
	Posição		2
	Comprimento		18
		HVI	16
		HVII	1
		HVI + HVII	1
		HVIII	3
	Posição + Comprimento		3

Relativamente às heteroplasmas de posição, foram detetadas 5 heteroplasmas em 5 das sequências (6%). Em uma das amostras, na posição 16129, foi visível a sobreposição de uma adenina (A) e de uma guanina (G). Em duas das sequências, na posição 16192, observou-se a presença de uma citosina (C) e de uma timina (T). Foi também identificada, na posição 16286, a sobreposição de uma citosina (C) e de uma timina (T). Na posição 204 verificou-se a presença de uma citosina (C) e de uma timina (T). De acordo com o código IUPAC, as heteroplasmas foram classificadas, respetivamente, como 16129R, 16192Y, 16286Y e 204Y.

Em relação às heteroplasmas de comprimento, foram identificadas 21 sequências com heteroplasma (25.3%) (tabela 4.6). No segmento de HVI compreendido entre as posições 16182 e 16192, observou-se a presença de heteroplasma em 16 das sequências analisadas (19.3%). Entre as posições 303 a 316, na região HVII, apenas 1 heteroplasma foi detetada (1.2%). Em 3 das sequências constatou-se a presença de heteroplasma no segmento de HVIII compreendido entre as posições 567 e 574 (3.6%). Observou-se ainda, em 1 amostra (1.2%), a presença, em simultâneo, de heteroplasma na região HVI e HVII da região controlo total do DNAm.

4.3. Estudo dos Haplogrupos

A partir das 83 amostras analisadas identificaram-se 74 haplótipos diferentes, que foram inseridos, por sua vez, em 50 haplogrupos distintos. A classificação das sequências nos respetivos haplogrupos foi conseguida através de uma árvore filogenética, desenhada por van Over e Kayser em 2009 [52], a *PhyloTree*, que inclui os polimorfismos específicos de todos os haplogrupos e subhaplogrupos existentes. Para que um haplótipo seja incorporado corretamente num determinado haplogrupo é necessário que apresente, pelo menos, as alterações específicas que caracterizem esse haplogrupo. Como tal, dois indivíduos com diferenças nos seus haplótipos podem ser incluídos no mesmo haplogrupo. Por exemplo, dois dos indivíduos analisados neste estudo caracterizam-se, respetivamente, pelos haplótipos 16185T **16209C** 16223T 16311C **16320T** 16327T 73G 150T 185A 189G 263G 315.1C e 16185T 16223T 16311C **16317G** 16327T 73G 150T 185A 189G **200G** 263G **309.1C** 315.1C. Embora sejam divergentes em relação aos polimorfismos assinalados, ambos pertencem ao mesmo haplogrupo, ao L3e1a1, uma vez que apresentam as alterações específicas que o definem: 16185T, 16311C e 185A.

Através da análise dos resultados obtidos, constatou-se que 67 dos indivíduos (81%) foram agrupados no macrohaplogrupo L, característico de populações africanas, nomeadamente de linhagens de África subsariana [52,69]. Esta elevada percentagem de haplogrupos tipicamente africanos já foi descrita por diversos autores, em estudos com diferentes populações do continente africano [106,108,109]. Pereira e Salas, em estudos efetuados à população nativa de Moçambique, obtiveram um padrão genético de haplótipos característico apenas de comunidades africanas [55,56].

O conjunto de haplogrupos determinado neste estudo mostrou uma expressão bastante diversificada relativamente aos haplogrupos inseridos no macrohaplogrupo L, sendo os haplogrupos L2a (26.5%), L0a (18.1%), L3e1 (8.4%), L0d (7.2%), L1c (3.6%), L3e3 (3.6%), L3d (2.4%) e L3e1a (2.4%), os que apresentam uma dominância maior. Pereira, em 2001, alcançou resultados semelhantes aos do presente estudo, tendo como haplogrupos mais frequentes, os haplogrupos L2a (43.1%), L0a (14.7%), L3e1 (8.3%), L0d (7.3%), L1c (4.6%), L3b (3.7%) e L3e1a (3.7%) [56]. Também Salas, em 2002, obteve resultados equivalentes com supremacia dos haplogrupos L2a (29.3%), L0a (28.3%), L3d (6.5%), L3e1 (5.9%), L1c (5.5%), L0d (4.2%) e L3e3 (3.9%) [55]. Comparando os resultados do presente estudo com os de Pereira e Salas, podemos verificar que a composição genética da população

imigrante moçambicana é representativa da população nativa, uma vez que apresenta um padrão genético mitocondrial semelhante ao do território de origem.

A presença de haplogrupos típicos do continente africano na população de Lisboa, como resultado da recente imigração, pode conduzir à alteração da composição genética mitocondrial da população portuguesa, com consequente aumento da variabilidade genética.

No presente estudo, verificou-se ainda que os restantes 16 indivíduos (19%) foram inseridos em haplogrupos caraterísticos de populações euroasiáticas, entre os quais os haplogrupos H, J, K, M, R, T e U. A reduzida frequência destes haplogrupos revela a pequena contribuição de linhagens femininas europeias e asiáticas na composição genética da população em estudo. A sua presença na população imigrante pode ser consequência de recente fluxo de genes ou refletir o passado histórico de migrações de indivíduos destas regiões.

A figura 4.1 resume detalhadamente o conjunto de haplogrupos obtidos para a população imigrante moçambicana.

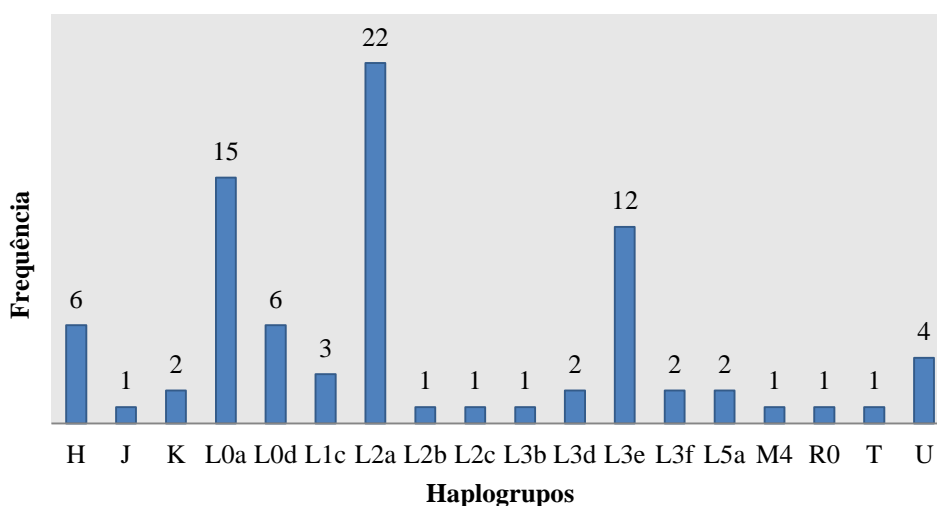


Figura 4.1 – Representação dos diferentes haplogrupos e da respetiva frequência relativos aos 83 indivíduos da população imigrante moçambicana.

4.3.1. Haplogrupo L0

A primeira ramificação do macrohaplogrupo L ocorreu há aproximadamente 140-160 mil anos, com a separação do haplogrupo L0 [112]. Este haplogrupo encontra-se predominantemente nas regiões Este, Sudeste e Sul africanas [69,113] e pode dividir-se em vários subhaplogrupos – L0a, L0b, L0d, L0f e L0k [52]. Neste estudo, apenas os subhaplogrupos L0a e L0d foram identificados. No conjunto dos 83 indivíduos analisados, 15 enquadravam-se no subhaplogrupo L0a (18.1%) e 6 no subhaplogrupo L0d (7.2%). Detalhadamente e relativamente ao subhaplogrupo L0a, verificou-se que 8 das amostras inseriram-se no subhaplogrupo L0a1 (9.7%) e 7 no subhaplogrupo L0a2 (8.4%). Relativamente ao subhaplogrupo L0d, 3 amostras identificaram-se como pertencentes ao subhaplogrupo L0d1 (3.6%) e 3 ao L0d2 (3.6%). O subhaplogrupo L0d foi o primeiro subhaplogrupo a divergir, há cerca de 100 mil anos, sendo a sua distribuição praticamente restrita ao povo Khoisan do Sul do continente africano, nomeadamente em Moçambique, justificando a sua presença nas amostras analisadas, e em populações da Tanzânia e Angola, onde é visível em altas frequências [112–116]. O subhaplogrupo L0a encontra-se vastamente disperso por toda a África subsariana e é considerado um marcador do povo Bantu [104,113]. Teve origem na região Este do continente africano, há cerca de

40-55 mil anos e atualmente é muito comum no Este, Centro e Sudeste de África e praticamente ausente nas restantes regiões africanas, justificando assim a alta frequência deste subhaplogrupo neste estudo [55,69,104,112].

4.3.2. Haplogrupo L1

O haplogrupo L1 divergiu, dos haplogrupos L2 a L6, há aproximadamente 140-150 mil anos. Subdivide-se nos subhaplogrupos L1b e L1c, cujas ramificações datam, respetivamente, de há cerca de 30 e 80 mil anos, e tem origem em regiões centrais e ocidentais do continente africano [69,112,113, 115,117]. Dos 83 indivíduos, apenas 3 foram classificados como pertencentes ao haplogrupo L1 (3.6%). Estas 3 amostras foram inseridas no subhaplogrupo L1c, mais precisamente uma amostra em L1c1 e duas em L1c3. Nenhuma foi enquadrada no subhaplogrupo L1b. A baixa percentagem do haplogrupo L1 nas amostras dos imigrantes moçambicanos é consistente com a sua origem geográfica e a localização atual deste haplogrupo, estando praticamente ausente no Este e Sudeste africano [55,69]. O subhaplogrupo L1c, também de origem central e ocidental, é predominante nestas regiões e é considerado um marcador da população indígena Bantu. Em muito baixa frequência, pode ser encontrado em regiões do Sudeste e Sudoeste africanos, como resultado da migração do povo Bantu [104,112,118,119]. Este subhaplogrupo, pode ser também encontrado em indivíduos Afro-americanos, nomeadamente da parte Sul da América, do qual o Brasil faz parte, e que aí se estabeleceram após comércio transatlântico de escravos [55,118,120].

4.3.3. Haplogrupo L2

O haplogrupo L2, juntamente com o haplogrupo L3, constitui 70% de todos os haplogrupos de DNAm encontrados em África subsariana [112,121]. Este haplogrupo é muito frequente em populações do Oeste africano, seu local de origem, e encontra-se amplamente disperso por todo o continente africano, nomeadamente em comunidades do Sudeste e Centro-Oeste africano, como consequência das expansões Bantu, e em comunidades do Noroeste, devido ao Comércio transaariano de escravos [122–124]. A sua formação data de há aproximadamente 90-105 mil anos e é composto por 5 subhaplogrupos – L2a, L2b, L2c, L2d e L2e [52,112]. Neste estudo, foram identificados os subhaplogrupos L2a, L2b e L2c. Dos 83 indivíduos analisados, 22 inseriram-se no subhaplogrupo L2a (26.5%), o mais frequente do presente estudo, 1 no subhaplogrupo L2b (1.2%) e 1 no L2c (1.2%). Detalhadamente e relativamente ao subhaplogrupo L2a, verificou-se que 21 das amostras enquadravam-se no subhaplogrupo L2a1 (25.3%) e 1 no subhaplogrupo L2a5 (1.2%). Estes subhaplogrupos dominam a região Sudeste, atingindo alta frequência na população de Moçambique. Quanto ao subhaplogrupo L2a1, 10 indivíduos foram identificados como pertencentes ao subhaplogrupo L2a1a (12.1%), o subhaplogrupo com maior representação neste estudo, 9 classificados como L2a1b (10.8%) e 2 ficaram confinados ao próprio subhaplogrupo L2a1 (2.4%). A alta frequência do subhaplogrupo L2a1a no Sudeste africano, onde Moçambique está inserido, já foi relatada por vários autores [55,115,120] e vai de encontro aos resultados obtidos neste estudo. O haplogrupo L2a é o mais comum e o mais amplamente distribuído em África, sendo também muito frequente nos Afro-Americanos (19%), como consequência do comércio transatlântico de escravos que se estabeleceu entre a América e o Oeste africano [55,56,125,126]. A sua origem geográfica é difícil de determinar, uma vez que apresenta alta frequência e alta diversidade em todo o continente africano [56,124,125,127,128]. Há autores que consideram que este haplogrupo teve uma origem mais a Oeste [127], enquanto que outros consideram uma origem mais central, expandindo-se posteriormente e quase concomitantemente para Este e Oeste africano [112]. Este subhaplogrupo está associado às migrações do povo Bantu e estima-se que a sua origem foi há cerca de 45-55 mil anos [112,113]. Os subhaplogrupos L2b e L2c são dominantes e praticamente restritos ao Oeste e Centro-Oeste de África,

provavelmente os seus locais de origem [55,112]. São raros no Sudeste, o que pode justificar a sua baixa frequência nos indivíduos da população em estudo [56,121,129].

4.3.4. Haplogrupo L3

Com origem na região Este do continente africano, estima-se que o superhaplogrupo L3 tenha surgido há aproximadamente 60-75 mil anos [55,112]. Este superhaplogrupo é constituído por linhagens pertencentes a países de África subsariana bem como por linhagens euroasiáticas, uma vez que dá origem aos haplogrupos M e N, que posteriormente derivam para todos os restantes haplogrupos não L [112,130]. O haplogrupo L3 encontra-se vastamente disperso por África, especialmente em populações Bantu do Este, Sudeste e Oeste, e divide-se em vários subhaplogrupos, desde o L3a ao L3h [52,112,113]. De todas as amostras analisadas, 17 inseriram-se neste haplogrupo (20.5%). Dos subhaplogrupos identificados, o que apresenta uma maior frequência é o subhaplogrupo L3e com 12 indivíduos (14.5%). Os subhaplogrupos L3d e L3f foram observados, mas com apenas dois indivíduos cada (2.4%, total 4.8%). Com 1 indivíduo (1.2%), identificou-se o subhaplogrupo L3b. Este último subhaplogrupo atinge a sua maior incidência em populações Bantu do Oeste africano e a sua presença no Sudeste pode ser consequência da expansão desta comunidade e daí a sua baixa frequência em Moçambique, resultado consistente com Pereira [56]. O subhaplogrupo L3d, também característico do Oeste africano, tem maior expressão em países como Angola e Tanzânia, e pode ser ainda encontrado nos Afro-americanos. Salas, em 2002, reportou que alguns subtipos do subhaplogrupo L3d podem ser observados no Sudeste africano [55]. O subtipo L3d1 foi um dos subtipos identificados neste estudo. O subhaplogrupo L3e é o mais comum e antigo subhaplogrupo de L3, com uma idade estimada em 45 mil anos [55]. Postula-se que este subhaplogrupo surgiu no Centro de África, atual região do Sudão, a partir da qual se dispersou um pouco por toda a África subsariana, atingindo o Sudeste e, por consequência Moçambique, através das comunidades Bantu, que aí se estabeleceram [112,120]. Este subhaplogrupo também é visualizado, com alta frequência, em linhagens provenientes do Brasil [55]. O L3e1 foi o subtipo predominante no conjunto das amostras analisadas, com 7 indivíduos (8.4%), e é o subconjunto identificado com maior frequência na região Sudeste, nomeadamente em Moçambique [55,112]. Este fato é consistente com os resultados obtidos no estudo realizado por Pereira à população nativa de Moçambique [56]. Os restantes subhaplogrupos, L3e2, L3e3 e L3e5, com um total de 5 indivíduos, também foram observados. Os subtipos L3e2 e L3e5 são mais comuns em países do Centro-Oeste e o subtipo L3e3 mais frequente no Oeste [55,112]. Relativamente ao subhaplogrupo L3f, ele tem origem no Este do continente africano, sendo praticamente restrito a esta região. A sua distribuição geográfica pode ser um marcador do fluxo transatlântico de escravos, uma vez que este subhaplogrupo encontra-se bem representado em indivíduos Afro-americanos [55,104,131]. O seu aparecimento em indivíduos moçambicanos pode indicar recente fluxo de genes.

4.3.5. Haplogrupo L5

Este haplogrupo, antigamente definido por Pereira como L1e, é raro (4-6%) e praticamente restrito ao Este africano, nomeadamente a países como o Egito, Sudão, Etiópia, Quênia, Ruanda e Tanzânia [56,69,112,120]. É um dos haplogrupos mais antigo, com um tempo de divergência estimado em aproximadamente 110 mil anos e divide-se nos subhaplogrupos L5a e L5b [52,117,132]. Pode também ser considerado um importante marcador do povo Bantu [112]. Duas das amostras analisadas foram identificadas como L5, mais precisamente como L5a. Uma amostra foi inserida no subhaplogrupo L5a1 e a outra no L5a2. Nenhum indivíduo foi enquadrado no subhaplogrupo L5b. Apesar do haplogrupo L5 ser muito raro em regiões fora do Este africano, devido a fluxos insignificantes de

genes, o subhaplogrupo L5a2, pode ser observado com alguma frequência em populações indígenas Bantu do Sudeste [112,120,133].

4.3.6. Haplogrupos euroasiáticos

Os haplogrupos H, I, J, K, M, T, U, V, W e X englobam quase todos os haplogrupos de DNAMt presentes nas populações europeias [134]. Neste estudo, haplogrupos como o H, J, K, M, T, U e ainda o R foram identificados nas amostras analisadas.

O haplogrupo U, embora específico de populações europeias, tem sido identificado também em indianos e comunidades africanas, nomeadamente do Norte de África [134–136]. Estima-se que este haplogrupo surgiu há mais de 50 mil anos, sendo dividido em vários subhaplogrupos, desde U1 até U9 [52,134,137]. Dos seus subhaplogrupos, apenas o U2, U5 e U6 foram observados no conjunto de amostras estudadas, num total de 4 (4.1%). Especificamente, os subhaplogrupos identificados foram o U2e1, U5a1, U5b1 e U6a, representados, cada um, por um indivíduo. O cluster U2, raro na Europa, é dominante na Índia. O cluster U5 é o mais frequente e mais antigo dos subhaplogrupos que derivam do haplogrupo U e, para além da Europa, pode ser observado esporadicamente no Próximo Oriente [137,138]. Contrariamente, o U6 é característico de populações do Norte e Noroeste africano. A presença destes subhaplogrupos noutras regiões do planeta pode ser um sinal de recente fluxo de genes [139].

O haplogrupo H é o haplogrupo europeu mais comum que apresenta uma distribuição geográfica muito vasta neste e noutros continentes. De fato, este haplogrupo é muito prevalente em praticamente todas as populações Europeias, nomeadamente no Oeste europeu, e é também muito comum no Norte de África, no Médio e no Próximo Oriente. Pode ser ainda encontrado, embora em menor frequência, no Nordeste da Índia e na Ásia Central [137,140,141]. Pensa-se que este haplogrupo teve origem no Médio Oriente, há 30-35 mil anos, tendo-se expandido posteriormente para a Europa. Devido à sua alta frequência e vasta distribuição está envolvido em recentes fenómenos de migração e fluxo de genes [140]. Identificaram-se, neste estudo, 6 indivíduos (7.2%) pertencentes a este haplogrupo. Especificamente, foram observados os subhaplogrupos H1 e H3, com 4 e 1 indivíduos, respetivamente. O sexto indivíduo foi classificado como pertencendo ao haplogrupo H visto não conter as alterações características de nenhum subhaplogrupo. Os subhaplogrupos H1 e H3 apresentam um padrão similar de origem europeia, nomeadamente na Europa Ocidental [142], sendo o H1 o mais frequente. O subhaplogrupo H1 tem sido observado no continente europeu, no Próximo Oriente, no Cáucaso, em Regiões do Norte de África e mais recentemente em comunidades do Sahara Central, possivelmente devido a expansões que se realizaram a partir da Península Ibérica [143–145]. O subhaplogrupo H3 tem uma localização mais restrita às regiões ocidentais europeias [142]. O aparecimento do haplogrupo H nos indivíduos imigrantes sugere miscigenações entre o povo europeu e moçambicano.

O haplogrupo K, também comum por toda a Europa, é considerado um subgrupo do haplogrupo U [146], que se subdivide nos subhaplogrupos K1, K2 e K3. Nas amostras analisadas, identificaram-se 2 que se enquadram nos subhaplogrupos K1 e K2. O subhaplogrupo K1, o mais comum, é particularmente frequente nas áreas do Mediterrâneo e pensa-se que tenha alcançado o continente europeu por migrações provenientes do Próximo Oriente [147,148].

Os restantes quatro indivíduos estudados foram classificados como pertencendo aos haplogrupos J, M, R, T, mais precisamente aos subhaplogrupos, J1c, M4a, R0 e T2b. J1c é encontrado em primeira instância na Europa, especialmente no Centro da Europa, nos Balcãs e na Ucrânia [149]. M4a é um subhaplogrupo específico indiano e visível em muito baixas frequências. R0 está presente sobretudo

no Oeste europeu, Próximo Oriente, Cáucaso e no Centro asiático [150]. T2c teve origem ao nível do Próximo Oriente e dispersou-se mais recentemente para a Europa [149].

4.4. Análise Filogenética

Para estimar a variabilidade genética entre diferentes populações recorreu-se à análise molecular da variância (AMOVA). Através dos resultados obtidos e de outros previamente publicados, realizou-se um estudo comparativo entre a população imigrante moçambicana, outras populações de origem africana e a população da região Sul de Portugal. As populações selecionadas e utilizadas na comparação filogenética estão ilustradas na tabela 4.7. O reduzido número de estudos efetuados com a totalidade da região controlo do DNAm_t justifica a comparação com poucas populações.

Tabela 4.7 – Populações selecionadas da literatura para estudos comparativos filogenéticos, tendo por base a região controlo total do DNAm_t.

Populações	Tamanho da amostra	Referências
Imigrantes moçambicanos	83	Presente estudo
Imigrantes angolanos	173	[109]
Imigrante cabo-verdianos	103	[108]
Imigrantes guineenses	80	[151]
Somália	190	[107]
Gana	193	[106]
Portugal	100	[152]

Para cada par de populações determinaram-se os valores de F_{st} e os respetivos valores de p , apresentados na tabela 4.8. Os valores das distâncias genéticas (F_{st}) entre as diferentes populações, permitiu elaborar uma representação gráfica onde foram visíveis as diferenças genéticas entre as populações (figura 4.2). Valores de p iguais ou superiores a 0,05 indicam semelhança entre duas populações. Verificou-se que os valores de p obtidos eram inferiores ao valor indicado, demonstrando que a população em estudo apresenta diferenças significativas para as populações selecionadas da literatura.

Tabela 4.8 – Valores de F_{st} e p relativos às populações do estudo filogenético obtidos através do programa *Arlequin*.

Populações	Imigrantes moçambicanos	Imigrantes angolanos	Imigrante cabo-verdianos	Imigrantes guineenses	Somália	Gana	Portugal
Imigrantes moçambicanos	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Imigrantes angolanos	0.01273	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Imigrante cabo-verdianos	0.06223	0.04388	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Imigrantes guineenses	0.04113	0.03422	0.00050	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Somália	0.02327	0.03572	0.06419	0.04452	-	0.00000±0.0000	0.00000±0.0000
Gana	0.04070	0.02606	0.01912	0.01055	0.05610	-	0.00000±0.0000
Portugal	0.11815	0.13053	0.19029	0.16957	0.08812	0.16977	-

A partir da figura 4.2 foi possível verificar a formação de dois grandes grupos: um composto pelas populações do continente africano e outro, mais distante, composto pela população portuguesa. Este resultado era expectável, uma vez que, em termos genéticos, as populações africanas são mais semelhantes entre si do que relativamente a populações de outros continentes.

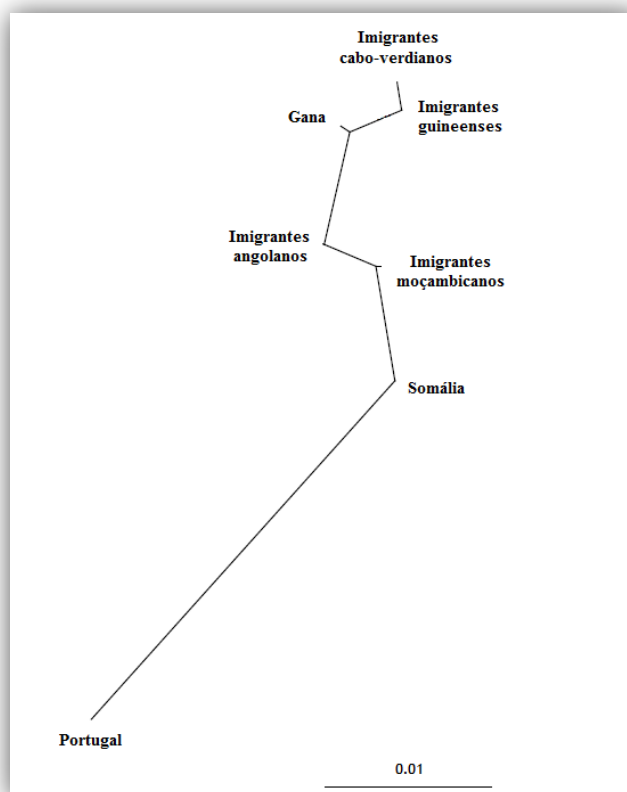


Figura 4.2 – Representação gráfica das distâncias genéticas entre a população imigrante moçambicana e as restantes seleccionadas da literatura.

Através da análise da matriz de distâncias interpopulacionais observou-se que a população de imigrantes moçambicanos a residir atualmente em Lisboa encontra-se geneticamente mais próxima da população de imigrantes angolanos residentes em Lisboa ($F_{st} = 0.01273$; $p=0.00000$). Alves em estudos realizados à comunidade Bantu, de origem angolana e moçambicana, confirmou a proximidade genética entre estes dois povos [153]. Este autor, nos seus trabalhos com outros marcadores genéticos, como UEPs (*Unique Event Polymorphisms*) e STRs, identificou altos níveis de homogeneidade genética entre os povos de Angola e Moçambique. Também Semo, em estudos efetuados com STRs, alcançou os mesmos resultados relativamente à proximidade entre os indivíduos angolanos e moçambicanos [154].

A população da Somália, a seguir à população de imigrantes angolanos, foi a população que apresentou menor distância genética para a população de imigrantes oriundos de Moçambique ($F_{st} = 0.02327$; $p=0.00000$). Semo, nos seus estudos, alcançou os mesmos resultados relativamente à proximidade da população moçambicana com a população da Somália [154].

A análise da matriz também evidenciou, por outro lado, que a população estudada se encontra geneticamente mais distante da população do Gana ($F_{st} = 0.04070$; $p=0.00000$), seguida da população de imigrantes da Guiné-Bissau ($F_{st} = 0.04113$; $p=0.00000$) e da população de imigrantes de Cabo Verde ($F_{st} = 0.06223$; $p=0.00000$). O afastamento da população estudada para a população do Gana, bem como, a proximidade desta última comunidade para a população da Guiné-Bissau já foi identificada

em estudos realizados com DNAmT [155]. O Gana é um país da costa ocidental africana que, após a colonização pelos portugueses, no século XV, estabeleceu comércio de mercadorias e escravos com outros países do continente, sendo o seu produto genético o conjunto de migrações provenientes de regiões do Norte e Nordeste, no século IX e X, bem como de migrações que ocorreram na periferia do Rio Volta. Atualmente, o principal habitante do Gana é o povo Akan, um importante grupo étnico do Oeste africano, fato que faz com que este país seja mais próximo de países da região Oeste, como Guiné-Bissau e Cabo Verde, e mais distante da região Este, do qual Moçambique faz parte [156]. O distanciamento dos imigrantes moçambicanos para a comunidade da Guiné-Bissau e de Cabo Verde já foi reportada em vários estudos com DNAmT [128,157] e com outros marcadores genéticos, como os InDels (*Insertion/Deletion*) [158]. Os principais grupos étnicos observados nos dias de hoje na Guiné-Bissau já habitavam o país aquando da chegada dos portugueses no século XV. A composição genética dos seus ocupantes foi influenciada por eventos demográficos que decorreram desde a pré-história, como movimentos migratórios, comércio de escravos e invasões consecutivas [157,159]. Este conjunto de situações é responsável pela distância para a população em estudo. Cabo Verde, por sua vez, teve uma comunicação limitada e dificultada com países do continente africano, por se tratar de um país isolado. Culturalmente, este país é o resultado de elementos europeus e africanos, nomeadamente da Guiné-Bissau, cuja mistura originou os designados crioulos ou mulatos. A proximidade geográfica entre Cabo Verde e Guiné-Bissau e a colonização do primeiro por indivíduos guineenses pode justificar a proximidade genética verificada entre os indivíduos dos dois países [108,160]. Ao invés, a grande distância genética verificada entre a população de imigrantes moçambicanos e a população de imigrantes cabo-verdianos pode ser justificada pelo fato da população de Cabo Verde apresentar uma composição genética própria.

Verificou-se ainda que a população mais distante da população imigrante estudada era a população portuguesa ($F_{st}=0.11815$; $p=0.00000$), como seria de esperar, por serem duas populações com características muito diferentes. Relativamente à população portuguesa, a população geneticamente mais próxima foi a da Somália ($F_{st}= 0.08812$; $p=0.00000$), da África oriental, e a mais distante foi Cabo Verde, da África ocidental.

Comparando as diferentes comunidades de imigrantes a viver em Lisboa, a que se encontra geneticamente mais próxima da população portuguesa é a moçambicana e a mais distante é a cabo-verdiana ($F_{st}= 0.19029$; $p=0.00000$), seguido da população da Guiné-Bissau ($F_{st}= 0.16957$; $p=0.00000$) e a de Angola ($F_{st}= 0.13053$; $p=0.00000$). Estes resultados estão de acordo com estudos previamente publicados relativos a outros marcadores genéticos, os InDels [158]. Pensa-se que estes resultados podem ser uma consequência da migração do povo Bantu que se estabeleceu em duas rotas principais ao longo da África subsariana: uma para Este e outra para Oeste. Estas duas rotas, ao evoluírem independentemente uma da outra, modularam distintamente os padrões genéticos das regiões Este e Oeste, uma vez que, este povo Bantu, ao fixar-se em diferentes regiões, substituiu parcialmente ou totalmente as populações indígenas Khoisan existentes [104,161]. Também os diferentes momentos de colonização nestas regiões, realizados pelo povo português, levaram a uma modulação diferencial dos padrões genéticos. Os movimentos migratórios entre diferentes países da Costa Oeste, onde Angola, Gana e Cabo Verde estão inseridos, pode ser também um dos fatores que contribuiu para estes países serem geneticamente mais próximos entre si e mais distantes dos países da Costa Este, onde Moçambique e Somália se localizam [162,163]. Um dos grandes marcos da história dos países da Costa Este que, por sua vez, também influenciou os padrões genéticos nestas regiões, foi a colonização Árabe [164,165]. Este povo é geneticamente mais próximo da população portuguesa do que das populações africanas [166–168], o que pode justificar o resultado das populações da Costa Este serem geneticamente mais próximas de Portugal do que as populações dos países da Costa Oeste.

5. Conclusões

A metodologia laboratorial utilizada no estudo do DNAm_t revelou ser adequada e eficiente para a realização deste trabalho, uma vez que permitiu a obtenção de sequências com elevada qualidade e com maior poder de discriminação, comparativamente com estudos efetuados apenas para as regiões HVI e HVII.

O conjunto de dados obtidos foi aceite com sucesso para integrar a base de dados internacional de DNAm_t mais conceituada no âmbito das Ciências Forenses, a EMPOP. Os resultados irão estar disponíveis para a comunidade científica com o código de acesso EMP00681.

Através da elevada percentagem de haplótipos únicos identificados na população imigrante moçambicana, num total de 68 haplótipos em 83 estudados (82%), demonstrou-se a grande variabilidade genética, já descrita por outros autores, para as populações africanas.

O haplótipo mais comum determinado neste estudo, foi identificado em 4 sequências e é constituído pelas seguintes alterações: 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL.

A análise das sequências de DNAm_t permitiu detetar um total de 1295 alterações nucleotídicas, que correspondem a 127 posições polimórficas. Verificou-se que as substituições foram a alteração mais frequente (77%) e dentro destas, as que tiveram uma incidência maior foram as transições entre pirimidinas (50%). Este elevado número de alterações nucleotídicas confirma dados anteriores relativos à grande heterogeneidade genética observada em populações africanas.

Foi na região HVII que se verificou a alteração mais frequente do estudo: a inserção de uma citosina na posição 315 (315.1C), com uma taxa de 100%. Este resultado vem confirmar estudos anteriormente realizados para várias populações, incluindo as africanas [73]. Outras alterações detetadas e também comuns foram os polimorfismos 263G (93%), 16223T (81%) e 16519C (60%).

As alterações 16187T 16188G 16189C, identificadas em 17% das sequências, constituem o conjunto dos polimorfismos mais frequente que se observou no presente estudo.

O menor número de alterações polimórficas identificadas numa sequência relativamente à rCRS foi 5, enquanto que o maior número de polimorfismos determinados foi de 30. O número médio de alterações nucleotídicas por amostra encontradas neste estudo foi de 16.

Com os valores da diversidade de sequência e nucleotídica obtidos, 0,9962 e 0,0152, respetivamente, foi possível, mais uma vez, confirmar a elevada diversidade genética desta população de imigrantes, que é característica das populações do continente africano.

A taxa de heteroplasmia total determinada neste estudo foi de 28%. Relativamente às heteroplasmias de posição, 5 foram detetadas e classificadas, de acordo com o código IUPAC, em 16129R, 16192Y (duas), 16286Y e 204Y. Um total de 22 heteroplasmias de comprimento foram observadas, com a maioria dos casos a ocorrer no segmento HVI (19%).

A grande parte dos haplótipos determinados neste estudo, sensivelmente 81%, enquadrou-se em haplogrupos característicos africanos e típicos das regiões subsarianas. Os mais abundantes foram os haplogrupos L0, L2 e L3, sendo o L2 o haplogrupo que apresentou maior expressão, contabilizando

cerca de 29% de todas as amostras. Estes resultados estão de acordo com estudos realizados anteriormente para a população de Moçambique [55,56].

Uma pequena percentagem de haplogrupos identificados (19%) foram inseridos em haplogrupos característicos de populações europeias e asiáticas, realçando a baixa contribuição destas populações no padrão genético da população imigrante moçambicana.

A análise filogenética permitiu verificar a formação de dois grandes grupos: um composto pelas populações do continente africano e outro, mais distante, composto pela população portuguesa. Através do cálculo das distâncias genéticas interpopulacionais foi possível observar que a população de imigrantes moçambicanos se encontra geneticamente mais próxima da população de imigrantes angolanos a residir em Lisboa ($F_{st} = 0.01273$). Contrariamente, e como seria de esperar, a população mais distante foi a população portuguesa ($F_{st} = 0.11815$).

Relativamente às diferentes comunidades de imigrantes que residem atualmente em Lisboa, verificou-se que a população em estudo foi a que apresentou menor distância genética da população portuguesa e a mais distante foi a cabo-verdiana. No que diz respeito à população portuguesa, a população geneticamente mais próxima foi a da Somália, da parte oriental de África, e a mais distante foi a de Cabo Verde, da África ocidental.

O presente estudo revela uma recente introdução, na população de Lisboa, de linhagens características de comunidades africanas, que poderão levar à modulação dos padrões genéticos nesta região e, consequentemente, em Portugal, se ocorrerem miscigenações entre indivíduos da população imigrante moçambicana e indivíduos de Lisboa.

Este trabalho demonstrou, assim, a importância da realização de estudos populacionais e da análise do DNAm, uma vez que, com este marcador genético, é possível traçar a história e fenómenos migratórios de populações, que ocorreram no passado, bem como aferir relações filogenéticas entre diferentes comunidades.

6. Estudos Futuros

Futuramente, seria interessante estudar a região codificante do DNAm para retirar conclusões mais detalhadas relativamente aos haplótipos de cada indivíduo e para confirmar os haplogrupos correspondentes a cada haplótipo. O objetivo seria aumentar a qualidade e rigor do estudo e verificar se haveria alterações na composição genética da população imigrante moçambicana, isto é, se haveria maior poder de discriminação com o estudo de todo o DNAm, do que apenas com o estudo da região controlo total do DNAm.

7. Referências

- [1] J. Cummins, "Mitochondrial DNA in mammalian reproduction," *Rev. Reprod.*, vol. 3, pp. 172–82, 1998.
- [2] F. Legros, F. Malka, P. Frachon, et al., "Organization and dynamics of human mitochondrial DNA," *J. Cell Sci.*, vol. 117, pp. 2653–2662, 2004.
- [3] R. E. Jensen, A. E. Aiken Hobbs, K. L. Cervený, et al., "Yeast mitochondrial dynamics: Fusion, division, segregation, and shape," *Microsc. Res. Tech.*, vol. 51, pp. 573–583, 2000.
- [4] M. van der Laan, S. E. Horvath, and N. Pfanner, "Mitochondrial contact site and cristae organizing system," *Curr. Opin. Cell Biol.*, vol. 41, pp. 33–42, 2016.
- [5] B. Alberts, A. Johnson, J. Lewis, et al., "Molecular Biology of the Cell," *Garland Science*, 5th ed., 2008.
- [6] M. M. Nass, and S. Nass, "INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS.I. Fixation and Electron Staining Reactions," *J Cell Biol.*, pp.593–611, 1963.
- [7] J.-W. Taanman, "The mitochondrial genome: structure, transcription, translation and replication," *Biochim. Biophys. Acta - Bioenerg.*, vol. 1410, pp. 103–123, 1999.
- [8] J. W. Ballard and M. C. Whitlock, "The incomplete natural history of mitochondria," *Mol. Ecol.*, vol. 13, pp. 729–744, 2004.
- [9] M. Lembring, "Application of Mitochondrial DNA Analysis in Contemporary and Historical Samples," *Uppsala Universitet*, vol. 62, 2013.
- [10] M. Tanaka, and T. Ozawa, "Strand Asymmetry in Human Mitochondrial DNA Mutations," *Genomics*, vol. 22, pp. 327–335, 1994.
- [11] S. Nonin-Lecomte, F. Dardel, and P. Lestienne, "Self-organisation of an oligodeoxynucleotide containing the G- and C-rich stretches of the direct repeats of the human mitochondrial DNA," *Biochimie*, vol. 87, pp. 725–735, 2005.
- [12] S. Lutz, H. J. Weisser, J. Heizmann, et al., "Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany," *Int. J. Legal Med.*, vol. 111, pp. 67–77, 1998.
- [13] M. Ingman, H. Kaessmann, S. Pääbo, et al., "Mitochondrial genome variation and the origin of modern humans," *Nature*, vol. 408, pp. 708–713, 2000.
- [14] S. Lutz, H. Wittig, H. J. Weisser, et al., "Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions?," *Forensic Sci. Int.*, vol. 113, pp. 97–101, 2000.
- [15] B. A. Malyarchuk, I. B. Rogozin, V. B. Berikoy, et al., "Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region," *Hum. Genet.*, vol. 111, pp. 46–53, 2002.
- [16] S. Anderson, A. T. Bankier, B. G. Barrell, et al., "Sequence and organization of the human mitochondrial genome," *Nature*, vol. 290, pp. 457–465, 1981.
- [17] R. M. Andrews, I. Kubacka, P. F. Chinnery, et al., "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA," *Nat. Genet.*, vol. 23, p. 147, 1999.

- [18] H.-J. Bandelt, A. Kloss-Brandstätter, M. B. Richards, et al., "The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies," *J. Hum. Genet.*, vol. 59, pp. 66–77, 2014.
- [19] D. Arora, A. Singh, V. Sharma, et al., "HgsDb: Haplogroups Database to understand migration and molecular risk assessment," *Bioinformation*, vol. 11, pp. 272–275, 2015.
- [20] B. Budowle, M. W. Allard, M. R. Wilson, et al., "FORENSICS AND MITOCHONDRIAL DNA: Applications, Debates, and Foundations," *Annu. Rev. Genomics Hum. Genet.*, vol. 4, pp. 119–141, 2003.
- [21] H. H. Cells, D. Bogenhagen, and D. A. Clayton, "The number of Mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells," *J. Biol. Chem.*, vol. 249, pp. 7991–7995, 1974.
- [22] P. Fernández-Silva, J. A. Enriquez, and J. Montoya "Replication and transcription of mammalian mitochondrial DNA," *Exp. Physiol.*, pp. 41–56, 2003.
- [23] R. E. Giles, H. Blanc, H. M. Cann, et al., "Maternal inheritance of human mitochondrial DNA," *Proc. Natl. Acad. Sci.*, vol. 77, p. 6715–6719, 1980.
- [24] P. Sutovsky, R. D. Moreno, J. Ramalho-Santos, et al., "Ubiquitin tag for sperm mitochondria," *Nature*, vol. 402, pp. 371–372, 1999.
- [25] W. Parson, T.J. Parsons, R. Scheithauer, et al., "Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence," *Int J Legal Med*, vol. 111, pp. 124–132, 1998.
- [26] W. M. Brown, M. J. George, and A. C. Wilson, "Rapid evolution of animal mitochondrial DNA," *P.N.a.S.*, vol. 76, pp. 1967–1971, 1979.
- [27] B. Pakendorf, and M. Stoneking, "Mitochondrial Dna and Human Evolution," *Annu. Rev. Genomics Hum. Genet.*, vol. 6, pp. 165–183, 2005.
- [28] A. A. Johnson, and K. A. Johnson, "Fidelity of Nucleotide Incorporation by Human Mitochondrial DNA Polymerase," *J Biol Chem.*, vol. 276, 2001.
- [29] P. Awadalla, A. Eyre-Walker, and J. M. Smith, "Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA," *Science*, vol. 286, pp. 2524–2525, 1999.
- [30] A. Eyre-Walker, N. H. Smith, and J. M. Smith, "How clonal are human mitochondria?," *Proc. Biol. Sci.*, vol. 266, pp. 477–83, 1999.
- [31] E. Hagelberg, N. Goldman, P. Lió, et al., "Evidence for mitochondrial DNA recombination in a human population of island Melanesia," *Proc. Biol. Sci.*, vol. 266, pp. 485–492, 1999.
- [32] M. Stoneking, "DNA and Recent Human Evolution," *Evol. Anthropol.*, vol. 2, pp. 60–73, 1989.
- [33] V. Macaulay, M. Richards, and B. Sykes, "Mitochondrial DNA recombination - no need to panic," *P Roy Soc L. B Bio*, vol. 266, pp. 2037–2039, 1999.
- [34] T. Kivisild, and R. Villems, "Questioning evidence for recombination in human mitochondrial DNA," *Science*, vol. 288, p. 1931, 2000.
- [35] R. N. Lightowlers, P. F. Chinnery, D. M. Turnbull, et al., "Mammalian mitochondrial genetics: Heredity, heteroplasmy and disease," *Trends Genet.*, vol. 13, pp. 450–455, 1997.
- [36] E. Hagström, C. Freyer, B. J. Battersby, et al., "No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline," *Nucleic Acids Res.*, vol. 42,

pp. 1111–1116, 2014.

- [37] A. Brandstätter, and W. Parson, “Mitochondrial DNA heteroplasmy or artefacts—a matter of the amplification strategy?,” *Int. J. Legal Med.*, vol. 117, pp. 180–184, 2003.
- [38] A. Alonso, A. Salas, C. Albarrán, et al., “Results of the 1999-2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the GEP-ISFG: An inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor,” *Forensic Sci. Int.*, vol. 125, pp. 1–7, 2002.
- [39] A. Ramos, C. Santos, L. Mateiu, et al., “Frequency and Pattern of Heteroplasmy in the Complete Human Mitochondrial Genome,” *PLoS One*, vol. 8, pp. 19–21, 2013.
- [40] M. Lagerström-Fermér, C. Olsson, L. Forsgren, et al., “Heteroplasmy of the human mtDNA control region remains constant during life,” *Am. J. Hum. Genet.*, vol. 68, pp. 1299–1301, 2001.
- [41] K. A. Roberts, and C. Calloway, “Characterization of Mitochondrial DNA Sequence Heteroplasmy in Blood Tissue and Hair as a Function of Hair Morphology,” *J. Forensic Sci.*, vol. 56, pp. 46–60, 2011.
- [42] A. M. Spicer, T. J. Kun, B. N. Sacks, et al., “Mitochondrial DNA sequence heteroplasmy levels in domestic dog hair,” *Forensic Sci. Int. Genet.*, vol. 11, pp. 7–12, 2014.
- [43] J. Naue, S. Hörer, T. Sängner, et al., “Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA,” *Mitochondrion*, vol. 20, pp. 82–94, 2015.
- [44] C. Herrnstadt, and N. Howell, “An evolutionary perspective on pathogenic mtDNA mutations: Haplogroup associations of clinical disorders,” *Mitochondrion*, vol. 4, pp. 791–798, 2004.
- [45] L. C. Greaves, A. K. Reeve, R. W. Taylor, et al., “Mitochondrial DNA and disease,” *J. Pathol.*, vol. 226, pp. 274–286, 2012.
- [46] L. A. Tully, T. J. Parsons, R. J. Steighner, et al., “A Sensitive Denaturing Gradient-Gel Electrophoresis Assay Reveals a High Frequency of Heteroplasmy in Hypervariable Region 1 of the Human mtDNA Control Region,” *Am. J. Hum. Genet.*, vol. 67, pp. 432–443, 2000.
- [47] C. D. Calloway, R. L. Reynolds, G. L. Herrin, et al., “The Frequency of Heteroplasmy in the HVII Region of mtDNA Differs across Tissue Types and Increases with Age,” *Am. J. Hum. Genet.*, vol. 66, pp. 1384–1397, 2000.
- [48] L. Prieto, M. Montesino, A. Rodríguez-Monge, et al., “Measuring by fragment analysis the proportion of length variants in samples carrying length heteroplasmy at the homopolymeric C-stretch in mitochondrial HVII region,” *Int. Congr. Ser.*, vol. 1261, pp. 103–105, 2004.
- [49] N. Sondheim, C. E. Glatz, J. E. Tirone, et al. “Neutral mitochondrial heteroplasmy and the influence of aging,” *Hum. Mol. Genet.*, vol. 20, pp. 1653–1659, 2011.
- [50] J. B. Stewart, and P. F. Chinnery, “The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease,” *Nat. Rev. Genet.*, vol. 16, pp. 530–542, 2015.
- [51] S. Lutz-Bonengel, T. Sängner, S. Pollak, et al., “Different methods to determine length heteroplasmy within the mitochondrial control region,” *Int. J. Legal Med.*, vol. 118, pp. 274–281, 2004.
- [52] M. van Oven, and M. Kayser, “Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation,” *Hum. Mutat.*, vol. 30, pp. 386–394, 2009.

- [53] W. M. Brown, "Polymorphisms of mitochondrial DNA of humans as revealed by restriction endonuclease analysis," *Proc. natn. Acad. Sci. U.S.A.*, vol. 77, pp. 3605–3609, 1980.
- [54] A. Nagai, I. Nakamura, F. Shiraki, et al., "Sequence polymorphism of mitochondrial DNA in Japanese individuals from Gifu Prefecture," *Leg. Med.*, pp. S292–S294, 2003.
- [55] A. Salas, M. Richards, T. De la Fe, et al., "The making of the African mtDNA landscape," *Am. J. Hum. Genet.*, vol. 71, pp. 1082–111, 2002.
- [56] L. Pereira, V. Macaulay, A. Torroni, et al., "Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade," *Ann. Hum. Genet.*, vol. 65, pp. 439–458, 2001.
- [57] H. Afonso Costa, M. Carvalho, V. Lopes, et al., "Mitochondrial DNA sequence analysis of a native Bolivian population," *J. Forensic Leg. Med.*, vol. 17, pp. 247–253, 2010.
- [58] D. Castro De Guerra, C. Figuera Pérez, C. M. Bravi, et al., "Sequence variation of mitochondrial DNA control region in North Central Venezuela," *Forensic Sci. Int. Genet.*, vol. 6, pp. 2011–2013, 2012.
- [59] Q. Mairal, C. Santos, M. Silva, et al., "Linguistic isolates in Portugal: Insights from the mitochondrial DNA pattern," *Forensic Sci. Int. Genet.*, vol. 7, pp. 618–623, 2013.
- [60] M. Knapp, K. A. Horsburgh, S. Prost, et al., "Complete mitochondrial DNA genome sequences from the first New Zealanders.," *Proc. Natl. Acad. Sci.*, vol. 109, pp. 18350–18354, 2012.
- [61] P. Soares, J. A. Trejaut, J. H. Loo, et al., "Climate change and postglacial human dispersals in Southeast Asia," *Mol. Biol. Evol.*, vol. 25, pp. 1209–1218, 2008.
- [62] R. K.-B. Atig, S. Hsouna, E. Beraud-Colomb, et al., "Mitochondrial DNA: properties and applications," *Arch. Inst. Pasteur Tunis*, vol. 86, pp. 3–14, 2009.
- [63] E. D. Ladoukakis, and E. Zouros, "Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions," *J. Biol. Res.*, vol. 24, pp. 1–7, 2017.
- [64] C. Stringer, "New perspectives on the Neanderthals," *Evol. Anthropol. Issues, News, Rev.*, vol. 11, pp. 58–59, 2003.
- [65] R. L. Cann, M. Stoneking, and A. C. Wilson, "Mitochondrial DNA and human evolution," *Nature*, vol. 325, pp. 31–36, 1987.
- [66] D. M. Behar, R. Villems, H. Soodyall, et al., "The Dawn of Human Matrilineal Diversity," *Am. J. Hum. Genet.*, vol. 82, pp. 1130–1140, 2008.
- [67] B. Harrub, and B. Thompson, "The Demise Of Mitochondrial Eve," *Homo*, pp. 1–7, 2010.
- [68] N. Howell, I. Kubacka, and D. A. Mackey, "How rapidly does the human mitochondrial genome evolve?," *Am. J. Hum. Genet.*, vol. 59, pp. 501–9, 1996.
- [69] M. K. Gonder, H. M. Mortensen, F. A. Reed, et al., "Whole-mtDNA genome sequence analysis of ancient african lineages," *Mol. Biol. Evol.*, vol. 24, pp. 757–768, 2007.
- [70] D. M. Behar, M. Van Oven, S. Rosset, et al., "A 'copernican' reassessment of the human mitochondrial DNA tree from its root," *Am. J. Hum. Genet.*, vol. 90, pp. 675–684, 2012.
- [71] T. Kivisild, "Maternal ancestry and population history from whole mitochondrial genomes," *Investig. Genet.*, vol. 6, p. 3, 2015.
- [72] Y.-S. Chen, A. Olckers, T. G. Schurr, et al., "mtDNA Variation in the South African Kung and

- Khwe - and Their Genetic Relationships to Other African Populations,” *Am. J. Hum. Genet.*, vol. 66, pp. 1362–1383, 2000.
- [73] B. Budowle, M. R. Wilson, J. A. DiZinno, et al., “Mitochondrial DNA regions HVI and HVII population data,” *Forensic Sci. Int.*, vol. 103, pp. 23–35, 1999.
 - [74] S. Lutz, H.-J. Weisser, J. Heizmann, et al., “mtDNA as a tool for identification of human remains Identification using mtDNA,” *Int. J. Legal Med.*, vol. 109, pp. 205–209, 1996.
 - [75] J. E. Cooper, and M. E. Cooper “Wildlife Forensic Investigation: Principles and Practice,” *CRC Press*, 2013
 - [76] T. C. Boles, C. C. Snow, and E. Stover, “Forensic DNA testing on skeletal remains from mass graves: a pilot project in Guatemala,” *J. Forensic Sci.*, vol. 40, pp. 349–55, 1995.
 - [77] A. Hernandez, I. Mamely, J. J. Sanchez, et al., “Mitochondrial DNA analysis of ancient human teeth from a XVIth century archeological excavation,” *Int. Congr. Ser.*, vol. 1239, pp. 601–604, 2003.
 - [78] W. Goodwin, A. Linacre, and P. Vanezis, “The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims,” *Electrophoresis*, vol. 20, pp. 1707–1711, 1999.
 - [79] M. Nilsson, "Mitochondrial DNA in Sensitive Forensic Analysis," Uppala Universitet, 2007.
 - [80] P.L. Ivanov, M. J. Wadhams, R. K. Roby, et al., "Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II,” *Nat Genet.*, vol. 12, pp. 417–20, 1996.
 - [81] M. D. Coble, “The identification of the Romanovs: Can we (finally) put the controversies to rest?,” *Investig. Genet.*, vol. 2, p. 20, 2011.
 - [82] J. M. Butler, ”Advanced Topics in Forensic DNA Typing,” *Academic Press*, 2011.
 - [83] A. Röhl, B. Brinkmann, L. Forster, et al., “An annotated mtDNA database,” *Int. J. Legal Med.*, vol. 115, pp. 29–39, 2001.
 - [84] W. Parson, and A. Dür, “EMPOP-A forensic mtDNA database,” *Forensic Sci. Int. Genet.*, vol. 1, pp. 88–92, 2007.
 - [85] E. Ruiz-Pesini, M. T. Lott, V. Procaccio, et al., “An enhanced MITOMAP with a global mtDNA mutational phylogeny,” *Nucleic Acids Res.*, vol. 35, pp. 823–828, 2007.
 - [86] M. Ingman, and U. Gyllensten, “mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences,” *Nucleic Acids Res.*, vol. 34, pp. D749–D751, 2006.
 - [87] Y. S. Lee, W.-Y. Kim, M. Ji, et al., “MitoVariome: a variome database of human mitochondrial DNA,” *BMC Genomics*, vol. 10, p. S12, 2009.
 - [88] M. Attimonelli, N. Altamura, R. Benne, et al., “MitBASE: A comprehensive and integrated mitochondrial DNA database,” *Nucleic Acids Res.*, vol. 27, pp. 128–133, 1999.
 - [89] J. Damas, J. Carneiro, A. Amorim, et al., “MitoBreak: The mitochondrial DNA breakpoints database,” *Nucleic Acids Res.*, vol. 42, pp. 1261–1268, 2014.
 - [90] B. Zimmermann, A. W. Röck, A. Dür, et al., “Improved visibility of character conflicts in quasi-median networks with the EMPop NETWORK software,” *Croat. Med. J.*, vol. 55, pp. 115–120, 2014.

- [91] L. Prieto, B. Zimmermann, A. Goios, et al., "The GHEP-EMPOP collaboration on mtDNA population data - A new resource for forensic casework," *Forensic Sci. Int. Genet.*, vol. 5, pp. 146–151, 2011.
- [92] C. Rego, M. F. Mendes, J. Rebelo, et al., "Perfil dos imigrantes em Portugal : por países de origem e regiões de destino," *Revista Portuguesa de Estudos Regionais*, vol. 24, 2010.
- [93] B. P. Mcevoy, J. E. Powell, M. E. Goddard, et al., "Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs," *Genome Research*, vol. 21, pp. 821–829, 2011.
- [94] PORDATA: Base de Dados Portugal Contemporâneo, 2017. (online: <https://www.pordata.pt>)
- [95] P. S. Walsh, D. a. Metzger, and R. Higuchi, "Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material," *Biotechniques*, vol. 54, pp. 506–513, 1991.
- [96] K. Mullis, F. Faloona, S. Scharf, et al., "Specific enzymatic amplification of dna in vitro: the polymerase chain reaction," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 51, pp. 263–273, 1986.
- [97] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Pnas*, vol. 74, pp. 5463–5467, 1977.
- [98] G. Tully, W. Bär, B. Brinkmann, et al., "Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles," *Forensic Sci. Int.*, vol. 124, pp. 83–91, 2001.
- [99] W. Parson, L. Gusmão, D. R. Hares, et al., "DNA Commission of the International Society for Forensic Genetics : Revised and extended guidelines for mitochondrial DNA typing Forensic Science International : Genetics DNA Commission of the International Society for Forensic Genetics," *Forensic Sci. Int. Genet.*, vol. 13, pp. 134–142, 2014.
- [100] L. Excoffier, and H. E. Lischer, "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows," *Molecular Ecology Resources*, vol. 10, pp. 564–567, 2010.
- [101] J. Felsenstein, "PHYLIP (Phylogeny Inference Package) version 3.695." Distributed by the author, *Department of Genome Sciences*, University of Washington, Seattle, 2013.
- [102] R. D. M. Page, "TreeView: an application to display phylogenetic trees on personal computers," *Comput Appl Biosci*, vol. 12, pp. 357–358, 1996.
- [103] A. Salas, T. De La Fé, B. Sobrino, et al., "Mitochondrial DNA variability patterns in Southeast Africa and forensic implications," *Int. Congr. Ser.*, vol. 1239, pp. 541–545, 2003.
- [104] S. Beleza, L. Gusmão, A. Amorim, et al., "The genetic legacy of western Bantu migrations," *Hum. Genet.*, vol. 117, pp. 366–375, 2005.
- [105] M. Carvalho, V. Bogas, F. Balsa, et al., "Mitochondrial control region data of 3 ethnic groups from angola," *Forensic Sci. Int. Genet. Suppl. Ser.*, vol. 3, pp. 125–126, 2011.
- [106] L. Fendt, A. Röck, B. Zimmermann, et al., "MtDNA diversity of Ghana: A forensic and phylogeographic view," *Forensic Sci. Int. Genet.*, vol. 6, pp. 244–249, 2012.
- [107] M. Mikkelsen, L. Fendt, A. W. Röck, et al., "Forensic and phylogeographic characterisation of mtDNA lineages from Somalia," *Int. J. Legal Med.*, vol. 126, pp. 573–579, 2012.
- [108] P. Morais, A. Amorim, C. V. Da Silva, et al., "Genetic portrait of Lisboa immigrant population

- from Cabo Verde with mitochondrial DNA analysis,” *J. Genet.*, vol. 94, pp. 509–512, 2015.
- [109] F. Simão, H. A. Costa, C. V. Da Silva, et al., “Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA,” *Forensic Sci. Int. Genet.*, vol. 15, pp. 33–38, 2014.
 - [110] M. Carvalho, C. Mendes, H. Antunes, et al., “mtDNA analysis in Portuguese populations (Central Portugal and Azores Islands): Polymorphic sites in control region sequences,” *Int. Congr. Ser.*, vol. 1239, pp. 535–539, 2003.
 - [111] A. Carracedo, W. Bär, P. J. Lincoln, et al., “DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing,” *Forensic Science International*, vol. 79, pp. 121–125, 2000.
 - [112] A. Rosa, and A. Brehm, “African human mtDNA phylogeography at-a-glance,” *J. Anthropol. Sci.*, vol. 89, pp. 25–58, 2011.
 - [113] Q. D. Atkinson, R. D. Gray, and A. J. Drummond, “Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa,” *Proc. R. Soc. B Biol. Sci.*, vol. 276, pp. 367–373, 2009.
 - [114] C. Barbieri, M. Vicente, J. Rocha, et al., “Ancient substructure in early mtDNA lineages of Southern Africa,” *Am. J. Hum. Genet.*, vol. 92, pp. 285–292, 2013.
 - [115] T. Kivisild, M. Reidla, E. Metspalu, et al., “Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears,” *Am. J. Hum. Genet.*, vol. 75, pp. 752–770, 2004.
 - [116] C. M. Schlebusch, M. Lombard, and H. Soodyall, “MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa,” *BMC Evol. Biol.*, vol. 13, p. 56, 2013.
 - [117] T. Rito, M. B. Richards, V. Fernandes, et al., “The first modern human dispersals across Africa,” *PLoS One*, vol. 8, pp. 1–16, 2013.
 - [118] C. Batini, V. Coia, C. Battaglia, et al., “Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa,” *Mol. Phylogenet. Evol.*, vol. 43, pp. 635–644, 2007.
 - [119] L. Quintana-Murci, H. Quach, C. Harmant, et al., “Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, pp. 1596–1601, 2008.
 - [120] S. Plaza, A. Salas, F. Calafell, et al., “Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola,” *Hum. Genet.*, vol. 115, pp. 439–447, 2004.
 - [121] E. Watson, P. Forster, M. Richards, et al., “Mitochondrial footprints of human expansions in Africa,” *Am. J. Hum. Genet.*, vol. 61, pp. 691–704, 1997.
 - [122] S. A. Tishkoff, M. K. Gonder, B. M. Henn, et al., “History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation,” *Mol. Biol. Evol.*, vol. 24, pp. 2180–2195, 2007.
 - [123] N. Harich, M. D. Costa, V. Fernandes, et al., “The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages,” *BMC Evol Biol*, vol. 10, p. 138, 2010.
 - [124] M. Silva, F. Alshamali, P. Silva, et al., “60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2,” *Sci. Rep.*, vol. 5,

p. 12526, 2015.

- [125] M. W. Allard, D. Polanskey, K. Miller, et al., "Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set," *Forensic Sci. Int.*, vol. 148, pp. 169–179, 2005.
- [126] D. C. Johnson, S. Shrestha, H. W. Wiener, et al., "Mitochondrial DNA diversity in the African American population," *HHS Public Access*, vol. 26, pp. 445–451, 2015.
- [127] A. Salas, M. Richards, M.-V. Lareu, et al., "The African diaspora: mitochondrial DNA and the Atlantic slave trade," *Am. J. Hum. Genet.*, vol. 74, pp. 454–465, 2004.
- [128] A. Rosa, A. Brehm, T. Kivisild, et al., "MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region," *Ann. Hum. Genet.*, vol. 68, pp. 340–352, 2004.
- [129] A. Torroni, C. Rengo, V. Guida, et al., "Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates?," *Am. J. Hum. Genet.*, vol. 69, pp. 1348–1356, 2001.
- [130] P. Soares, F. Alshamali, J. B. Pereira, et al., "The expansion of mtDNA haplogroup L3 within and out of Africa," *Mol. Biol. Evol.*, vol. 29, pp. 915–927, 2012.
- [131] V. Černý, V. Fernandes, M. D. Costa, et al., "Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup," *BMC Evol. Biol.*, vol. 9, p. 63, 2009.
- [132] C. Batini, J. Lopes, D. M. Behar, et al., "Insights into the demographic history of African pygmies from complete mitochondrial genomes," *Mol. Biol. Evol.*, vol. 28, pp. 1099–1110, 2011.
- [133] P. Shen, T. Lavi, T. Kivisild, et al., "Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation," *Hum. Mutat.*, vol. 24, pp. 248–260, 2004.
- [134] S. Finnilä, I. E. Hassinen, L. Ala-Kokko, et al., "Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis," *Am. J. Hum. Genet.*, vol. 66, pp. 1017–1026, 2000.
- [135] J. C. Rando, F. Pinto, A. M. Gonzalez, et al., "Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations," *Ann. Hum. Genet.*, vol. 62, pp. 531–550, 1998.
- [136] M. A. Bermisheva, I. A. Kutuev, T. Y. Korshunova, et al., "Phylogeographic analysis of mitochondrial DNA in the Nogays: A strong mixture of maternal lineages from Eastern and Western Eurasia," *Mol. Biol.*, vol. 38, pp. 516–523, 2004.
- [137] M. Richards, V. Macaulay, E. Hickey, et al., "Tracing european founder lineages in the near eastern mtDNA pool," *Am. J. Hum. Genet.*, vol. 67, pp. 1251–1276, 2000.
- [138] T. Kivisild, M. J. Bamshad, K. Kaldma, et al., "Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages," *Curr. Biol.*, vol. 9, pp. 1331–1334, 1999.
- [139] L. Pereira, N. M. Silva, R. Franco-Duarte, et al., "Population expansion in the North African Late Pleistocene signalled by mitochondrial DNA haplogroup U6," *BMC Evol. Biol.*, vol. 10, p. 390, 2010.
- [140] A. Achilli, C. Rengo, C. Magri, et al., "The molecular dissection of mtDNA haplogroup H

confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool,” *Am. J. Hum. Genet.*, vol. 75, pp. 910–8, 2004.

- [141] P. Brotherton, W. Haak, J. Templeton, et al., “Europe PMC Funders Group Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans,” *Europe PMC Funders Group*, vol.4, 2014.
- [142] L. Pereira, M. Richards, A. Goios, et al., “High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium,” *Genome Res.*, vol. 15, pp. 19–24, 2005.
- [143] U. Roostalu, I. Kutuev, E. L. Loogväli, et al., “Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: The Near Eastern and Caucasian perspective,” *Mol. Biol. Evol.*, vol. 24, pp. 436–448, 2007.
- [144] C. Ottoni, G. Primativo, B. H. Kashani, et al., “Mitochondrial haplogroup H1 in north Africa: An early Holocene arrival from Iberia,” *PLoS One*, vol. 5, pp. 1–7, 2010.
- [145] C. Turchi, F. Stanciu, G. Paselli, et al., “The mitochondrial DNA makeup of Romanians: A forensic mtDNA control region database and phylogenetic characterization,” *Forensic Sci. Int. Genet.*, vol. 24, pp. 136–142, 2016.
- [146] L. Simoni, F. Calafell, D. Pettener, et al., “mtDNA Haplogroups and Frequency Patterns in Europe,” *Am. J. Hum. Genet.*, vol. 66, pp. 1173–1177, 2005.
- [147] P. Endicott, J. J. Sanchez, E. Metspalu, et al., “The Unresolved Location of Ötzi’s mtDNA Within Haplogroup K,” *Am J Phys Anthropol*, vol. 132, pp. 590–991, 2007.
- [148] V. Coia, G. Cipollini, P. Anagnostou, et al., “Whole mitochondrial DNA sequencing in Alpine populations and the genetic history of the Neolithic Tyrolean Iceman,” *Sci. Rep.*, vol. 6, p. 18932, 2016.
- [149] M. Pala, A. Olivieri, A. Achilli, et al., “Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia,” *Am. J. Hum. Genet.*, vol. 90, pp. 915–924, 2012.
- [150] A. Brandstätter, B. Zimmermann, J. Wagner, et al., “Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East,” *BMC Evol. Biol.*, vol. 8, p. 191, 2008.
- [151] A. Amorim, J. Ribeiro, C. V. da Silva, et al., “Updated mtDNA study of Guinea-Bissau immigrant population living in Lisbon,” *Forensic Science International: Genetics Supplement Series*, 2017.
- [152] S. L. Marques, A. Goios, A. M. Rocha, et al., “Portuguese mitochondrial DNA genetic diversity - An update and a phylogenetic revision,” *Forensic Sci. Int. Genet.*, vol. 15, pp. 27–32, 2015.
- [153] I. Alves, M. Coelho, C. R. Gignoux, et al., “Genetic homogeneity across Bantu-speaking groups from Mozambique and Angola challenges early split scenarios between East and West Bantu populations,” *Hum. Biol.*, vol. 83, pp. 13–38, 2011.
- [154] A. C. Semo, M. Carvalho, A. Serra, et al., “Allelic frequencies of 15 autosomal STRs AmpF1 STR® Identifiler® Direct from two main population groups (Macua and Changana) in Mozambique,” 2013.
- [155] K. R. Veeramah, B. A. Connell, N. Pour, et al., “Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria,” *BMC Evol. Biol.*, vol. 10, p. 92, 2010.

- [156] A. A. Adeyemo, G. Chen, Y. Chen, et al., “Genetic structure in four West African population groups,” *BMC Genet.*, vol. 6, p. 38, 2005.
- [157] M. Carvalho, P. Brito, A. M. Bento, et al., “Paternal and maternal lineages in Guinea-Bissau population,” *Forensic Sci. Int. Genet.*, vol. 5, pp. 114–116, 2011.
- [158] A. Inácio, H. A. Costa, C. V. da Silva, et al., “Study of InDel genetic markers with forensic and ancestry informative interest in PALOP’s immigrant populations in Lisboa,” *Int. J. Legal Med.*, vol. 131, pp. 657–660, 2017.
- [159] A. Rosa, C. Ornelas, M. A. Jobling, et al., “Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective,” *BMC Evol. Biol.*, vol. 7, p. 124, 2007.
- [160] A. Amorim, R. Marques-Santos, C. Vieira-Silva, et al., “Genetic portrait of a native population of Cabo Verde living in Lisboa,” *Forensic Sci. Int. Genet.*, vol. 6, pp. 166–169, 2012.
- [161] G. Berniell-Lee, F. Calafell, E. Bosch, et al., “Genetic and demographic implications of the bantu expansion: Insights from human paternal lineages,” *Mol. Biol. Evol.*, vol. 26, pp. 1581–1589, 2009.
- [162] L. Pereira, L. Gusão, C. Alves, et al., “Bantu and European Y-lineages in Sub-Saharan Africa,” *Ann. Hum. Genet.*, vol. 66, pp. 369–378, 2002.
- [163] A. Brehm, L. Pereira, H.-J. Bandelt, et al., “Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade,” *Ann. Hum. Genet.*, vol. 66, pp. 49–60, 2002.
- [164] C. Serra, “Para a História da Arte Militar Moçambicana (1505-1920), ” *Cadernos tempo*, Maputo, 1983.
- [165] L. Bonate, “Islam in Northern Mozambique: A Historical Overview,” *Hist. Compass*, vol. 8, pp. 573–593, 2010.
- [166] K. K. Abu-Amro, J. M. Larruga, V. M. Cabrera, et al., “Mitochondrial DNA structure in the Arabian Peninsula,” *BMC Evol. Biol.*, vol. 8, p. 45, 2008.
- [167] V. Fernandes, F. Alshamali, M. Alves, et al., “The Arabian cradle: Mitochondrial relicts of the first steps along the Southern route out of Africa,” *Am. J. Hum. Genet.*, vol. 90, pp. 347–355, 2012.
- [168] J. L. Rodriguez-Flores, K. Fakhro, F. Agosto-Perez, et al., “Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations,” *Genome Res.*, vol. 26, pp. 151–162, 2016.

8. Anexos

8.1. Anexo A

Tabela 8.1 – Conjunto dos resultados obtidos para os 83 indivíduos imigrantes moçambicanos, quanto ao número de alterações polimórficas determinadas relativamente à rCRS, haplótipo e respetivo haplogrupo e codificação atribuída a cada sequência.

Amostra	Haplogrupo	Polimorfismos	Haplótipo																															
MOC_001	L3e1a1	12	16185T	16209C	16223T	16311C	16320T	16327T	73G	150T	185A	189G	263G	315.1C																				
MOC_002	L0a2a2	22	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL										
MOC_003	L0a1b1	24	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL								
MOC_005	L2a1a2	15	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	309.2C	315.1C																	
MOC_006	H1b	7	16189C	16356C	16519C	263G	309.1C	309.2C	315.1C																									
MOC_007	L3e1	8	16223T	16327T	73G	150T	189G	263G	309.1C	315.1C																								
MOC_008	L2a1a2	13	16223T	16278T	16286Y	16294T	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C																			
MOC_009	L2a1b1a	15	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C																	
MOC_010	L0a1a2	21	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	93G	185A	189G	200G	236C	247A	263G	315.1C	523DEL	524DEL											
MOC_011	U2e1	14	16051G	16129C	16183C	16189C	16362C	16519C	73G	152C	217C	263G	309.1C	315.1C	340T	508G																		
MOC_012	L1c3b1a	24	16129A	16163G	16187T	16189C	16209C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL								
MOC_013	H1b1+16362	9	16183C	16189C	16356C	16362C	16519C	263G	315.1C	523DEL	524DEL																							
MOC_014	L0a2a2	23	16148T	16172C	16173T	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL									
MOC_015	L0a1b1	24	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	204Y	236C	247A	263G	315.1C	523DEL	524DEL								
MOC_016	L3e1b2	10	16223T	16239T	16325DEL	73G	150T	185A	189G	263G	309.1C	315.1C																						
MOC_017	H1k	6	16189C	16290T	16519C	263G	309.1C	315.1C																										
MOC_018	L0a2a2	23	16093C	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL									
MOC_019	R0	5	16519C	263G	309.1C	309.2C	315.1C																											
MOC_020	L2a1b1a	15	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C																	
MOC_021	L3e1b2	11	16223T	16239T	16325DEL	73G	150T	185A	189G	194T	263G	309.1C	315.1C																					
MOC_022	L3b1a3	13	16124C	16223T	16278T	16293G	16311C	16362C	16519C	73G	263G	309.1C	315.1C	523DEL	524DEL																			
MOC_023	U6a	9	16172C	16219G	16278T	16357C	16519C	73G	263G	309.1C	315.1C																							
MOC_024	U5a1+a16192	9	16256T	16270T	16311C	16399G	73G	195C	200G	263G	315.1C																							
MOC_025	L0a2a2	23	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	150T	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL									
MOC_027	L3e1a1	13	16185T	16223T	16311C	16317G	16327T	73G	150T	185A	189G	200G	263G	309.1C	315.1C																			
MOC_028	L1c1	30	16117C	16129A	16172C	16173T	16188A	16189C	16223T	16256T	16278T	16291T	16293G	16294T	16311C	16360T	16368C	16519C	73G	151T	152C	182T	186A	195C	198T	247A	263G	297G	315.1C	316A	523DEL	524DEL		
MOC_029	L3d1a1a	12	16124C	16223T	16293G	16319A	73G	150T	152C	263G	309.1C	315.1C	523DEL	524DEL																				
MOC_030	L2c2	18	16223T	16264T	16278T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL														
MOC_031	L3e1	9	16223T	16327T	16343G	73G	150T	189G	200G	263G	315.1C																							
MOC_032	L0a1b1	25	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	375T	523DEL	524DEL							
MOC_033	L2a1b1a	17	16182C	16183C	16189C	16192T	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C															
MOC_034	L2a1b1a	16	16182C	16183C	16189C	16192T	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C																
MOC_035	L3d2b	10	16111T	16124C	16223T	73G	152C	199C	263G	315.1C	523DEL	524DEL																						
MOC_036	L2a1a2	14	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C																		
MOC_037	L5a1	23	16129A	16148T	16166G	16183DEL	16187T	16189C	16192T	16223T	16278T	16311C	16355T	16362C	73G	152C	182T	247A	263G	315.1C	444G	455.1T	459.1C	523DEL	524DEL									
MOC_038	L2a1a2	16	16223T	16278T	16286T	16291T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C	523DEL	524DEL																
MOC_039	L0d2c	20	16129A	16187T	16189C	16223T	16230G	16243C	16311C	16344T	16519C	73G	146C	152C	195C	247A	249DEL	294A	315.1C	523DEL	524DEL	498DEL												
MOC_041	H3c3	8	16519C	260A	263G	309.1C	309.2C	315.1C	524.1A	524.2C																								
MOC_042	L0a1b1	24	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL								
MOC_043	L3e2b	11	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																					
MOC_044	L2a1a1	19	16129A	16223T	16278T	16294T	16309G	16368C	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C	524.1A	524.2C	524.3A	524.4C													

Tabela 8.1 – (Continuação).

Amostra	Haplogrupo	Polimorfismos	Haplótipo																									
MOC_045	L0d1c	16	16187T	16189C	16223T	16230G	16234T	16243C	16311C	16519C	73G	146C	152C	195C	247A	315.1C	456T	498DEL										
MOC_046	L2b	19	16114A	16129A	16213A	16223T	16274A	16278T	16390A	73G	146C	150T	152C	182T	183G	195C	198T	204C	263G	309.1C	315.1C							
MOC_047	L0a2a2	22	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL				
MOC_048	K1a	10	16188T	16224C	16311C	73G	195C	263G	315.1C	497T	524.1A	524.2C																
MOC_049	L3e3	10	16093C	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																
MOC_050	L3f1b4a	11	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C															
MOC_051	L0a1b1	24	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL		
MOC_052	L0a2	22	16093C	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16362C	16519C	64T	93G	152C	189G	236C	247A	263G	315.1C	523DEL	524DEL				
MOC_053	H	5	16311C	16519C	150T	263G	315.1C																					
MOC_054	L2a1+16189(16192)	15	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C											
MOC_055	J1c2	13	16069T	16126C	16519C	73G	185A	188G	228A	263G	295T	309.1C	315.1C	462T	489C													
MOC_056	K2a	12	16224C	16278T	16311C	16519C	73G	146C	152C	263G	309.1C	315.1C	523DEL	524DEL														
MOC_057	L2a1b1a	15	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C											
MOC_058	L5a2	24	16129A	16148T	16166G	16183DEL	16187T	16189C	16192T	16223T	16278T	16311C	16355T	16362C	73G	152C	182T	247A	263G	315.1C	444G	455.1T	459.1C	523DEL	524DEL	527T		
MOC_059	M4a	13	16145A	16176T	16179T	16223T	16261T	16266T	16291T	16311C	16519C	73G	263G	315.1C	489C													
MOC_060	L2a1a2	13	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C													
MOC_061	L2a1a2	14	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C												
MOC_062	L2a1b1a	16	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C										
MOC_063	L0a2a2	22	16148T	16172C	16187T	16188A	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL				
MOC_064	L0a1b1	23	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	315.1C	523DEL	524DEL			
MOC_065	L3e1b2	11	16168T	16223T	16239T	16325DEL	73G	150T	185A	189G	263G	309.1C	315.1C															
MOC_068	L2a1a2	13	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C													
MOC_069	L0d2a1	21	16129A	16187T	16189C	16212G	16223T	16230G	16243C	16311C	16390A	16519C	73G	146C	152C	195C	198T	247A	309.1C	315.1C	498DEL	523DEL	524DEL					
MOC_070	L0a1b1	24	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL		
MOC_071	U5b1b1+@16192	7	16183C	16189C	16270T	73G	150T	263G	315.1C																			
MOC_072	L2a1b1a	17	16129A	16182C	16183C	16189C	16192Y	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C									
MOC_073	L0d1b2b2	23	16129A	16187T	16189C	16223T	16239T	16243C	16294T	16311C	16519C	73G	146C	152C	195C	247A	315.1C	498DEL	523DEL	524DEL	573.1C	573.2C	573.3C	573.4C	573.5C			
MOC_074	L3e3	10	16223T	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																
MOC_075	L0d2a	20	16129A	16187T	16189C	16212G	16223T	16230G	16243C	16311C	16390A	16519C	73G	146C	152C	195C	247A	309.1C	315.1C	498DEL	523DEL	524DEL						
MOC_076	T2b3+151	9	16126C	16294T	16296T	16304C	16519C	73G	151T	263G	315.1C																	
MOC_077	L2a1b1a	17	16182C	16183C	16189C	16192Y	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C									
MOC_078	L2a1b1a	17	16129R	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C									
MOC_079	L0d1b2b2	21	16129A	16187T	16189C	16223T	16239T	16243C	16294T	16311C	16519C	73G	146C	152C	195C	247A	315.1C	498DEL	523DEL	524DEL	573.1C	573.2C	573.3C					
MOC_080	L2a5	14	16223T	16224C	16256T	16278T	16309G	16390A	16519C	73G	146C	152C	182T	263G	315.1C	511T												
MOC_081	L2a1a3c	16	16093C	16223T	16256T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	309.1C	315.1C										
MOC_082	L3e3a	15	16223T	16265T	16519C	73G	150T	195C	263G	309.1C	315.1C	523DEL	524DEL	573.1C	573.2C	573.3C	573.4C											
MOC_083	L3f1b4a	11	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C															
MOC_084	L2a1a2	15	16092C	16223T	16278T	16286T	16294T	16309G	16355T	16390A	16519C	73G	146C	152C	195C	263G	315.1C											
MOC_085	L3e5	12	16041G	16223T	16355T	16519C	73G	150T	263G	315.1C	398C	480C	523DEL	524DEL														
MOC_086	H1+152	5	16519C	152C	263G	309.1C	315.1C																					
MOC_087	L1c3a	25	16129A	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	309.1C	315.1C	316A	408A	494T	523DEL	524DEL	
MOC_088	L2a1	13	16189C	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C													

